Entropy production and information processing in stochastic thermodynamics: Optimization, measurement, and erasure

Vegard Børve Sørdal

August 19, 2019

Thesis submitted for the degree of Philosophiæ Doctor

Department of Physics
University of Oslo

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 2166

ISSN 1501-7710

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Acknowledgement

I want to thank all my colleagues on the fourth floor, who have made these last four years a fantastic experience. Even though I have not directly worked with or collaborated with most of you, it has been a pleasure to work in close proximity with everyone.

This thesis would not exist if it were not for the help of Joakim Bergli and Yuri Galperin. Joakim is my main supervisor, and he is one of the most patient teachers I know. He never cuts any corners when discussing complex topics, and has been a great inspiration for how to think about analytical problem-solving. Yuri is my co-supervisor and his long experience and deep knowledge of physics, has been invaluable. I have been fortunate to have had advisors whose academical interests are closely aligned with my own, and together we have enjoyed countless exciting discussions. It has been a true pleasure.

I also want to thank Luiza Angheluta-Bauer for great collaboration and guidance while I was performing my teaching duties. Finally, to my family for always supporting me, and for watching Globus 2 with me when I was a child, which I suspect first sparked my interest in natural science.

Blindern, June 2019
Vegard Børve Sørdal

Contents

Chapter 1

Introduction

The everyday world is one of macroscopic variables. We talk about what the inner temperature of a perfect steak should be, the minimum pressure of a diving tank at which we should return to the surface, and how to optimally furnish a new apartment according to the available floor area. These macroscopic variables describe the gross state of the world; a coarse-grained interpretation of the universe. We never discuss what the exact configuration (position, velocity, interactions, etc.) of molecules that make up perfectly cooked steak should be. This exact configuration is called the microstate of the perfect steak. If we had a specific microstate that corresponds to a perfect steak, and then swapped around a few molecules we would get another, different, microstate. However, these two microstates would most certainly taste the same. The fact that we would not be able to distinguish these microstates by taste, is the reason why we care more about macroscopic states in an everyday description of the world. Macroscopic variables are those we can reliably measure and use to distinguish different systems from each other.

The macroscopic laws of thermodynamics were largely developed during the 18th and 19th centuries. These laws describe how macroscopic variables like temperature, pressure, and volume behave with respect to each other. As the foundation of thermodynamics, lie the four laws of thermodynamics, which describe how heat, energy, and entropy behave under various circumstances. Many famous statements, which even non-physicists are familiar with, come from these laws: *"energy can never be created nor destroyed, only change form"*, *"it's impossible to cool a system to absolute zero"*, and *"perpetual motion machines can not be created"*.

However, these laws were first postulated at a time where we did not know that the world was built up of elementary particles like atoms and electrons. They were formulated using macroscopic variables, which are just coarse-graining of the underlying microscopic variables. With the rise of statistical and quantum mechanics and a massive improvement in technological capabilities, we began to be able to detect and measure microstates directly. The natural question that arose was; How do we explain the empirically observed laws of thermodynamics, from the underlying microscopic behavior? In some cases, this was not too difficult. For example, the first law of thermodynamics, the conservation of energy, is deeply connected to time translation symmetry via Noether's theorem. In other cases, it was not so straightforward. The second law implies an arrow of time in physics, but how can the reversible microscopic dynamics of particles lead to irreversible macroscopic phenomena? One particular paradox which this thesis focuses on is Maxwell's demon. The resolution of this paradox revealed a deep connection between information and the laws of physics. This has had a large effect on physics, to such a degree that some researchers consider information to be the most fundamental constituent of the universe, rather than quarks or strings. Much of the work of this thesis is based on the relationship between information and thermodynamics, and how to optimize these information processing systems. We will discuss ideas from information theory, such as logical reversibility, measurement and erasure, what the equivalent physical processes of these somewhat abstract concepts are, and how they relate to the macroscopic laws of thermodynamics.

List of papers

This thesis is an article-based thesis, and as such the main text serves as an introduction to the minimal knowledge needed to read and understand the articles, which can be found at the very end of the thesis. What follows is a summary of all papers associated with this thesis.

1. Cooling by heating: Restoration of the third law of thermodynamics 2016
V.B. Sørdal, J. Bergli, Y.M. Galperin *Physical Review E 93 (3), 032102*
In this paper, we perform a detailed analysis of a quantum refrigerator powered by bosons. The refrigerator model appeared to violate the third law of thermodynamics, by allowing cooling to absolute zero in a finite amount of time. We show that the cooling power is exponentially quenched when the thermal energy approach

the scale of the energy level spacing.

2. Influence of measurement error on Maxwell's demon 2017
V.B. Sørdal, J. Bergli, Y.M. Galperin *Physical Review E 95 (6), 062129*
We show that errors in a symmetric binary measurement result in an error entropy S_ϵ, and that for optimal operation of a Szilard engine this error entropy dominates the total entropy production, even for very small measurement errors.

3. Quantum particle in a split box: Excitations to the ground state 2019
V.B. Sørdal, J. Bergli *Physical Review A 99 (2), 022121*
We introduce a method to achieve equal probability to find a quantum particle on either side of a barrier when it is inserted into a single-particle-box. By exciting only the first two energy levels, an asymmetric Szilard engine can reach the same efficiency as a symmetric one, without the need for information compression during erasure.

4. Deep reinforcement learning for robust quantum optimization 2019
V.B. Sørdal, J. Bergli *Submitted to Physical Review A*
We use deep reinforcement learning (DQL and DDPG), as well as traditional optimization techniques, to create robust protocols for the insertion of a potential barrier in an asymmetric quantum Szilard engine.

Structure of thesis

The structure of the thesis is as follows:

- Chapter 2 introduces the historical background to much of the work presented in the papers and serves as a motivation for the rest of the thesis. We introduce the laws of thermodynamics, and two systems that appeared to violate them. The second example, Maxwell's demon, is the motivating background for papers 2, 3 and 4.

- Chapter 3 covers basic concepts from statistical mechanics and information theory, which are needed as a foundation for all articles.

- Chapter 4 is an extension of Chapter 3, where we cover the quantum mechanical analogue for principles introduced in the previous chapter. This

chapter is especially important for papers 3 and 4, since those deal with quantum mechanical systems.

- Chapter 5 serves as an introduction to deep reinforcement learning, an exciting technique from the field of machine learning, which was employed for article number 4.
- Chapter 6 is a summary of unpublished work done in collaboration with the small biosystems lab at the University of Barcelona, where we study the energetics and entropy fluctuations of biomolecular motors.
- In the appendix, we include detailed calculations of an extension to paper 2. The calculations of all other papers are more or less contained in the papers themselves.
- Finally, all the papers are included in the last chapter.

Chapter 2

From steam machines to transistors

This chapter serves as a historical and philosophical introduction to controversies related to the second and third law of thermodynamics. In paper 1 we resolve an apparent violation of the third law of thermodynamics, and therefore we give a short brush up on its formulation and history in chapter 2.1. The three other articles in this thesis deal with questions related to another seeming violation, this time a violation of the second law. Hence, the larger part of this chapter is intended to introduce this apparent violation and discuss its three main proposed resolutions.

The laws of thermodynamics are among the most important laws in physics. Statistical mechanics is an essential tool in all fields of physics, and it has been adopted by many other sciences. Underlying the formulation of statistical mechanics is thermodynamics, the foundation of which is its main four laws. The zeroth law is quite straightforward; If two systems are in thermal equilibrium, call them A and B, and a third system, C, is in thermal equilibrium with B, A and C are also in equilibrium with each other. The first law is the conservation of energy; Any flow of energy through a system, in the form of heat or work, is compensated by a change its internal energy, such that the total change in energy of the universe is zero. The second and third laws are less straightforward, and will be discussed shortly.

2.1 Cooling to absolute zero

The third law of thermodynamics has its roots in the heat theorem, which was put forth by Walther Nernst in 1906 [1]. He stated that *"the entropy change in a chemical reaction tends to vanish as the temperature approaches absolute zero."*, and his work was expanded upon by Einstein [2] and Planck [3]. Einstein's statement of the third law is that the entropy of any substance tends to a constant value as the temperature falls to absolute zero

$$\lim_{T\to 0} S(T, X) = S_0. \tag{2.1}$$

Here X is any parameter of the system that its entropy may depend on. The third law of thermodynamics was formulated before quantum mechanics, yet it is really quantum mechanical in nature. A key property of quantum systems is that they have gapped energy spectrums; the possible energy-eigenstates are discretized and energies between these discrete states are unattainable. At zero temperature, the only energy state which the system can be in is its lowest energy state; the ground state. If this ground state has a degeneracy g, the indeterminacy of the exact eigenstate is g, and therefore the entropy is $S_0 = k_B \ln g$. Planck's statement is essentially the same, only he considered a perfect crystal, which has a non-degenerate ground state, and therefore the constant will be $S_0 = k_B \ln 1 = 0$.

A consequence of the third law of the thermodynamics is the unattainability principle; it is impossible to cool any system to absolute zero in finite time. This principle has been proved for many example systems, but a general proof does not yet exist. Without proof, the validity of the principle has to be checked on a case-to-case basis. One case that seemed to show a violation of the unattainability principle, arose in an article presenting a boson powered refrigerator [4]. The cooling power of the refrigerator was shown to scale linearly with temperature, which implies that absolute zero temperature can be reached in finite time. In paper 1, we resolve this apparent violation of the unattainability principle, by considering a fully quantum mechanical description of the system.

2.2 A violation of the second law?

The Industrial Revolution ($\sim 1750-1850$) was a period of monumental change in the United States and Europe. Automation of labor by machine tools, development

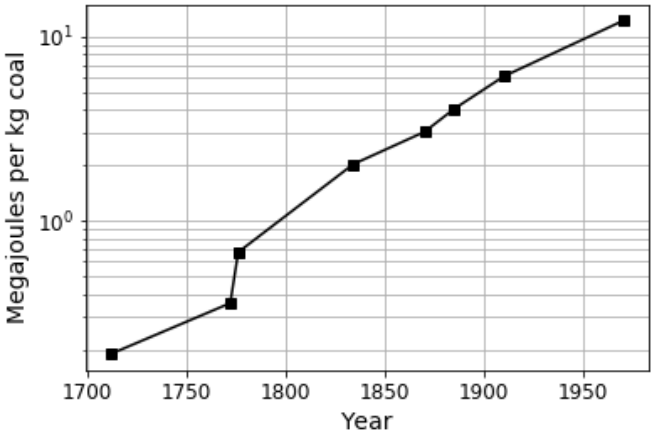

Figure 2.1: Graph showing the exponential increase in the efficiency of steam engines from 1700 to 1950. Data gathered by Dr. Grant Walker, University of Calgary.

of highly efficient factories, and centralization of production lines, dramatically changed society. Factories created new jobs and made old jobs obsolete. Farmers migrated in large numbers to the urban centers in search of work. One of the main driving forces of this revolution was the increased use of steam machines. Spurred on by the importance of these machines on the new society, science was also undergoing a sort of revolution, or more appropriately; a paradigm shift.

The first steam machines were primitive and inefficient, and to increase their efficiency, one first had to understand their driving force. And of course, steam machines are powered by the flow of heat. In the 18th century, the early days of the revolution, heat was believed to be a special kind of fluid. It was called *caloric*, and was a self-repellent weightless gas, that flowed from warmer to colder bodies [5]. However, during the early 19th century experiments were performed that disproved the caloric theory, and new research by the fathers of thermodynamics (Carnot, Joule, Clausius, Thomson, and others) discovered that heat was just another form of energy. They realized that motion and heat are mutually interchangeable; compressing a gas requires a certain amount of external energy (work), the compression heats up the gas so the work performed can be regained by allowing the gas to expand back to its original volume. How much of the work performed that can be regained by the expansion of the hot gas depends on how the exact design of the heat engine, and many new designs were introduced due to

the increased understanding of heat, like the Carnot-, Otto-, and Stirling engine. The exponential increase in the efficiency of steam engines from 1750 to 1950, can be seen in Fig. 2.1. Similarly to Moore's law, stating that the number of transistors per square inch on integrated circuits doubles every year, the efficiency of steam machines doubled every ~ 60 years over a period of 200 years.

A hot gas and a cold gas both consists of a collection of particles, the only difference is that the particles in the hot gas have a higher average kinetic energy than the ones in the cold gas. This connection between heat and the motion of particles gave rise to the new field of statistical mechanics, which can be considered the successor of thermodynamics. In statistical mechanics one studies how a systems macroscopic observables, like temperature and pressure of a gas, can be described as an ensemble average of its constituent microscopic properties. Peter G. Tait was a prominent scientist writing a book on statistical mechanics, and in 1867 he wrote a letter to James C. Maxwell, asking for hints on which topics to discuss in his book [6]. Maxwell's answer was the start of numerous debates and theoretical work on the relationship between physics and information theory, lasting 150 years until the present day.

2.2.1 Maxwell's demon

Maxwell's answer to Tait's letter was:

> *Any contribution I could make to that study is in the way of altering the point of view here and there for clearness or variety, and picking holes here and there to ensure strength and stability. (...) To pick a hole - say in the 2nd law of thermodynamics, that if two things are in contact the hotter cannot take heat from the colder without external agency. Now let A and B be two vessels (...)*

The thought-experiment that Maxwell went on to describe was the following: Imagine a box with two compartments, A and B, and a hatch between them as shown in Figure 1. The box contains an ideal gas at equilibrium temperature, and the temperature is the same in either compartment $T_A = T_B$. The velocities of the particles (indicated by the length of the arrows) in a gas at equilibrium are not all the same, but rather follows a distribution that Maxwell knew very well, since his name is in it; the Maxwell-Boltzmann velocity distribution, which is illustrated at the top of Fig. 2.2. Some particles move faster than others, and therefore have

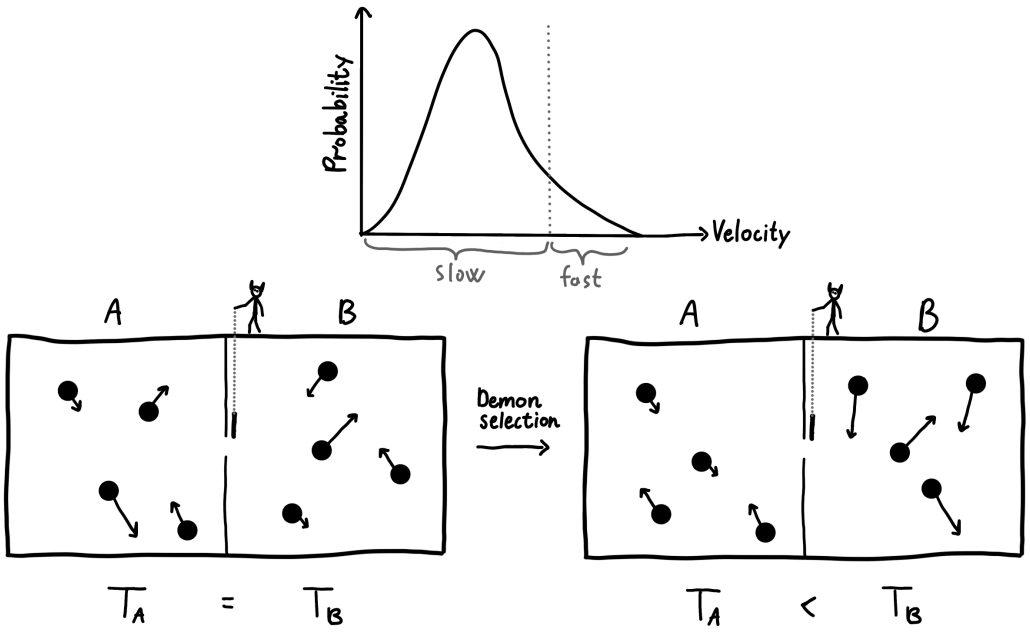

Figure 2.2: In the top of this figure we show a sketch of the Maxwell-Boltzmann distribution of an ideal gas. Below this, we illustrate the two compartments, A and B, and the demon lowering and raising a trapdoor to select which particles are allowed to pass between the chambers.

higher kinetic energy. For an ideal gas, temperature is just a thermodynamic average of the kinetic energy of all the particles. The task of the demon is to open the hatch and let "fast" particles (faster than the current average speed in the compartment) pass from compartment A to compartment B, while closing it to block the "slow" particles. Similarly, it opens and closes the hatch for slow and fast particles, respectively, moving from B to A. As the demon dutifully perform his task, over time the fast particles will gather in compartment B and the slow ones in compartment A. Since temperature is as previously mentioned related to the kinetic energy of the particles, what we (the demon to be precise) have achieved is to create a temperature difference between the compartments, where $T_B > T_A$. Assuming that the hatch is well oiled, so that the demon can move it without expending any energy, we have just violated the second law of thermodynamics. One of its many formulations is that heat can never flow from regions of low to high temperature without expending energy, which is exactly what the demon has just accomplished. If we now remove the hatch and put in a small turbine, high energy particles moving back from B to A would rotate the turbine and generate energy. After reaching equilibrium, where $T_A = T_B$, the state of the two compartments have returned to the initial state. The cycle is therefore a reversible process; a process that transforms a state back to itself without expending any energy. The demon can now go back to work and by continuously repeating the whole procedure we create infinite energy from nothing; a most severe violation of energy-conservation.

The thought-experiment was designed by Maxwell *"to show that the 2nd law of thermodynamics has only a statistical certainty"* [**?**]. If particle statistics could be influenced by an external agent (the demon) then the second law would not hold. Maxwell emphasized that the demon needed supernatural powers of observation and pinpoint precision. In his opinion, the operation was only a matter of scale, and we could in principle violate the second law, *"only we can't, not being clever enough"*.

2.2.2 Szilard engine

In 1929 Léo Szilard introduced a simplified version of Maxwell's demon, which is now known as the Szilard engine [7]. Maxwell's original thought-experiment consisted of a many-particle gas, but the same violation of the second law can be illustrated with a much simpler single-particle gas. The Szilard engine has

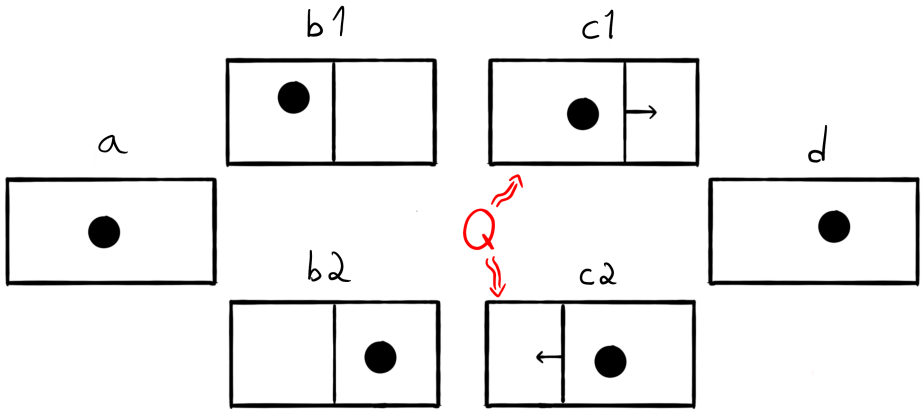

Figure 2.3: Schematic illustration of the Szilard engine protocol. When inserting a barrier in the center of a single-particle-box, the particle will either be confined to the right or to the left side with probability 1/2 each. After measuring which side the particle is found, we let the wall expand isothermally into the empty compartment work. During this isothermal expansion, work can be extracted from the gas pressure. After the wall has been entirely pushed to one side, we end up in a state identical to the initial one.

replaced Maxwell's demon as the standard second law violating information-thermodynamic thought-experiment, and has been the basis of most the theoretical work that followed Szilard, including this thesis.

Imagine a one-dimensional box, containing a single particle as shown in Fig. 2.3(a). Since the walls of the box are elastic, the particle does not lose any energy in collisions. The box is at equilibrium with a surrounding environment with a temperature T, and infinite heat-capacity. A barrier is inserted in the center of the box, such that the probability is 1/2 to find the particle on either side of it, as shown in Fig. 2.3(b1/b2). We now perform a measurement to determine which side of the box the particle is occupying. If it's found on the left side, we let single-particle gas expand isothermally by allowing the barrier to move into the empty compartment. When the particle collides with the barrier, energy is transferred from the particle to the barrier. The energy of the particle is replenished in the form of heat from the environment, as shown in Fig. 2.3(c1/c2). Once the barrier has been

moved entirely to one side of the box, such that the particle occupies the full volume again, we remove it. The final state of the box, Fig. 2.3(d), is now identical to the initial state: the particle occupies the full volume, and has a kinetic energy given by the temperature of the heat bath. Thus the process can be repeated. During the isothermal expansion, work is performed by the particle while moving the barrier (which could be extracted by e.g. attaching a pulley and weight to the barrier). The total work performed during the isothermal expansion is given by

$$W_{exp} = \int_{V/2}^{V} \frac{k_B T}{V}\, dV = k_B T \log 2, \tag{2.2}$$

and this work is gained by a full conversion from heat energy to work. Conservation of energy shows that, since the initial and final state is identical with the same energy, $W = -Q$. Similarly to Maxwell's demon, the Szilard engine extracts work from a single heat-bath with uniform temperature, in violation of the second law of thermodynamics. If the heat capacity of the environment was not infinite, the work extracted per cycle would be smaller than $k_B T \log 2$ since the final state would have lower energy than the initial one. Nevertheless, by continually repeating the process the heat bath would eventually be completely drained of energy, all of which would be converted to useful work. Szilard emphasized the necessity of performing a measurement for the engine to work. He connected the apparent violation of the second law with the state of the demon's knowledge, and believed that the resolution to the paradox was due to some hidden entropic cost associated with the measurement.

2.3 Three approaches

There are three main approaches to explain the apparent violation of the second law that Maxwell's demon and the Szilard engine implies.

1. The first approach focuses on the role of fluctuations, which are usually ignored in the idealized thought-experiments but will always be present in real systems.

2. The second approach follows Szilard's own belief and focuses on the entropic cost of performing measurements.

3. The third approach focuses on the fact that the demon has to store the information it obtains about the system. Unless this information is deleted, the

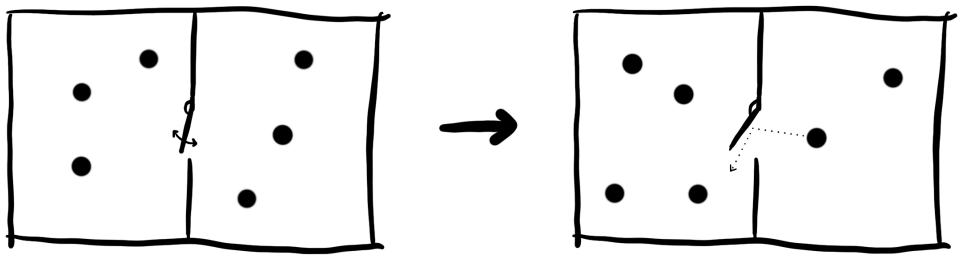

Figure 2.4: Illustration of Smoluchowski's spring-loaded trapdoor.

final state of the universe will not be identical to the initial state. The third approach focuses on the cost of deleting this memory

In the following section, we will go through each of these approaches.

2.3.1 Fluctuations

The first approach was to consider the effect of fluctuations. The first detailed analysis of the fluctuations in a Maxwell's demon-like system was done by Smoluchowski [8] in 1912. He removed the presence of an external observer, by replacing the demon with a trapdoor and a spring, as shown in Fig. 2.4. When relaxed, the spring keeps the trapdoor in the closed positing. The spring-loaded trapdoor allows particles to move from the right side into the left side, but blocks any particle moving in the opposite direction. After some period of time, particles will gather on the left side, building up a pressure difference between the two partitions. This pressure difference could be used to perform work, i.e. by replacing the trapdoor with a ratchet, lifting up a weight. The ultimate source of the work produced is the thermal energy of the gas, therefore this process seems to violate the second law of thermodynamics, just like the original Maxwell's demon. However, detailed analysis of the Smoluchowski trapdoor and similar apparatuses [8, 9, 10], all show that there is no true violation of the second law. Initially, the spring is at rest, with the trapdoor in the closed position. Particles hitting the trapdoor from the left side, bounces back and does not transfer any energy into the spring. Particles hitting the trapdoor from the right, push it open, and moves into the left partition. Since the spring is compressed when the trapdoor opens, it has to have a finite spring-constant. Therefore the spring itself constitutes a thermodynamic system, with potential and kinetic energy. Every time a particle hits

the trap door, energy is transferred into the spring, and it starts to oscillate about an equilibrium position. As Smoluchowski pointed out, if the impact of a single particle is enough to open the trapdoor, the trapdoor and spring have to be very light. This implies that after a few impacts with left-moving particles, the trapdoor would quickly start to randomly move between its opened and closed position, allowing particles on its left side move into the right side. Therefore there would not be a consistent buildup of particles on either side, and a pressure difference to extract work from would not develop.

On short time-scales, there could be small pressure differences between the partitions, which corresponds to a small decrease in entropy. However one of the key points of the second law, which is often forgotten, is that it is statistical in nature [11, 12]. This was pointed out by Maxwell already in 1878;

> *The truth of the second law is therefore a statistical, not a mathematical, truth, for it depends on the fact that the bodies we deal with consists of millions of molecules, and that we never can get a hold of a single molecule.*[6]

A proper definition of the second law of thermodynamics is that entropy can not spontaneously decrease when averaged over a suitable time-scale. What constitutes a suitable time-scale depends on the relaxation time of the system in question and is difficult to define in a general way. A quantitative description of the connection between fluctuations and the second law was given in 1993, when Evans et. al introduced the fluctuation theorem [13]. The theorem is actually a group of closely connected theorems [14, 15, 16], one of which relates the probability of observing a time-averaged entropy production of magnitude ΔS, to the probability of that it takes the opposite value $-\Delta S$

$$\frac{P(\Delta S)}{P(-\Delta S)} = e^{\Delta S/k_B} \geq 1 \tag{2.3}$$

Since the right side is always positive and larger than 1, the probability to observe fluctuations that temporarily "violate" the second law is always less or equal to the ones that obey it. Moreover, the relative probability of producing and consuming entropy increases exponentially with the amount of entropy change. Since entropy is extensive, the fluctuation theorem also shows that the probability to observe negative entropy fluctuations goes to zero for macroscopic systems.

The Smoluchowski trapdoor and similar apparatuses show that closed systems obeying purely Hamiltonian dynamics can not violate the second law of thermodynamics. Nevertheless, these thought-experiments take away the essential part of what constitutes Maxwell's demon: an external agent, collecting information about, and interacting with, the system. Smoluchowski himself allowed for the possibility of a modification of the second law, taking into account external agents:

> As far as we know today, there is no automatic, permanently effective perpetual motion machine, in spite of the molecular fluctuations, but such a device might, perhaps, function regularly if it were appropriately operated by intelligent beings. [8]

2.3.2 Measurement

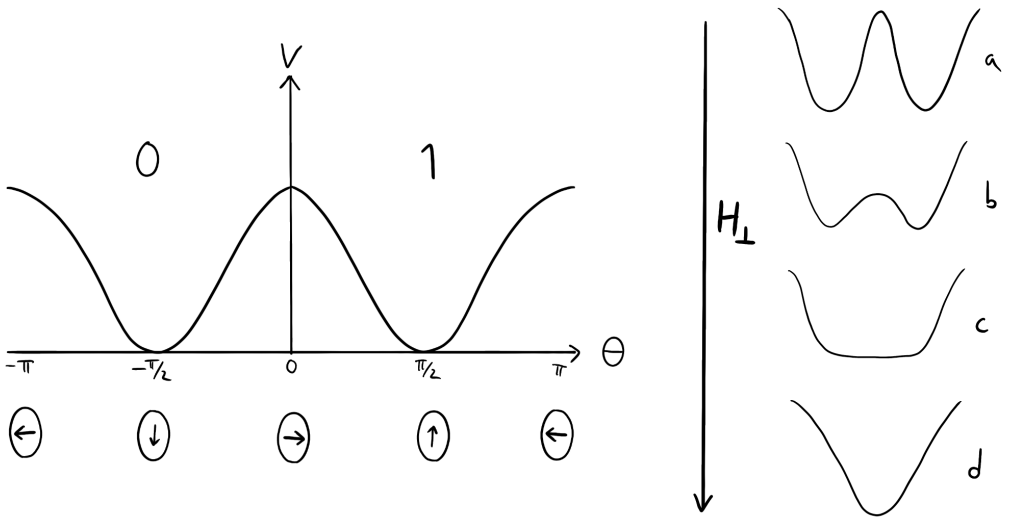

Figure 2.5: The bistable potential well of a single-domain ferromagnet, shown with the modulation of the potential under the application of a transverse magnetic field $H_\perp$. Illustration adapted from [17].

If one accepts the presence of the demon, and believe that the second law can not be violated, there must be some increase in entropy associated with one of the operations in the demon's cycle. The question that remained was then; which step in the cycle is responsible for the unaccounted entropy production? Szilard himself argued that the increase in entropy was due to the measurement carried

out by the demon. He first postulated that the second law has to be obeyed, and after eliminating all possible sources of entropy production (barrier insertion, expansion, and extraction), was left with the measurement process. In essence, his argument was that if the second law was to be obeyed, and the source of the extra entropy production needed to obey the second law could not be found via a statistical mechanical analysis of the operational processes in Szilard's engine; it is required that the work extracted is compensated by the entropic cost of measurement [18].

Szilard's view was later supported by Brillouin [19, 20], Gabor [21], and Rothstein [22]. Brillouin and Gabor presented specific models of dissipative measurements, sending light into the two compartment to see which side of the Szilard engine contains the particle after the barrier insertion. Light interacts with the particle and scatters if it is present, but in order to observe the scattered light, it has to be distinguishable from the background radiation. Since the electromagnetic field is in thermal equilibrium with the rest of the system, the blackbody radiation of the background has a mean energy of $k_B T$. Therefore, to distinguish the scattered photon from the background radiation it has to have an energy $h\nu \gg k_B T$. Using a photon with energy higher than the energy gained by the operation of Szilard's engine prevents any net extraction of work from the heat bath.

Efforts by Gabor and Brillouin to formulate a general theory on the entropic cost of measurement, based on their optical model, proved futile. A measurement can be described by an interaction between the system and a measurement apparatus, which results in a correlation between them. The state of the system can then be inferred by the state of the measurement apparatus. Of course, there exist measurement procedures to establish correlations between two systems, which also dissipate energy into the environment. The light scattering measurement is one of them, but there is no requirement that the measurement of a Maxwell demon state has to be performed by any kind of optical procedure. Their example is just one of many measurement procedures that dissipate energy. However, there is no general theory or principle showing that measurements are always accompanied by an entropic cost. On the contrary, examples of dissipation-free measurements have been given by Bennett [17, 23, 24, 25], one of which we will now discuss. Consider an ellipsoid piece of ferromagnetic material, so small that in the absence of a magnetic field consists of a single domain, magnetized in either parallel or anti-parallel to the ellipse axis. The potential landscape as a function of the angle

θ of the domain magnetization is illustrated in the left plot of Fig. 2.5.

We consider the anti-parallel magnetized state ($\theta = -\pi/2$) to be the logical state 0, while the parallel state ($\theta = \pi/2$) is the logical state 1. A longitudinal magnetic field $H_{\parallel}$ can be applied to bias the system in favor of either the parallel or anti-parallel state. The potential landscape can also be modulated by applying a transverse magnetic field $H_{\perp}$, as shown on the right side of Fig. 2.5(a-d), where we sketch the potential as a function of the applied field $H_{\perp}$. This modulation takes the system from a bistable potential as in Fig. 2.5(a), to a monostable potential as in Fig. 2.5(d). An intermediate "soft mode" occurs when the transverse field has reduced the central potential to zero, as shown in Fig. 2.5(c). Since the potential barrier between the 0 and 1 state is removed, the magnetization of a system in this state is very sensitive to applied longitudinal fields $H_{\parallel}$. This sensitivity allows us to reversibly copy information from one system to another. An example of such a measurement is illustrated in Fig. 2.6.

For any measurement to be reversible, the memory which we copy information into has to be in a standard reference state. Otherwise, a measurement would also erase information about what state the memory was initially in, and since this is a logically irreversible process it would generate additional entropy. We will discuss this in further detail in chapter 3.2.2. Therefore, the measurement apparatus consists of a reference bit in a known state (the 0 state in this case), a movable bit which starts in the same state as the reference bit, and a data bit, which is the state that we want to copy to the movable bit. As the movable bit enters the transverse magnetic field it is brought into the soft mode, where the bit becomes monostable. When the movable bit is brought out from the center of the transverse magnetic field, towards the data bit, it is very sensitive to the influence of longitudinal magnetic fields $H_{\parallel}$. The data bit, which is in the 1 state, exerts a small longitudinal filed in the direction of its magnetization, thereby biasing the movable bit towards the same magnetization as it is brought out of the transverse field. We assume the region of strong transverse field is wide enough so that by the time the movable bit reaches the bottom edge, the surrounding longitudinal filed is due entirely to the data bit, and has no influence from the reference bit further away. If the process of moving the bit from a reference to the data bit is performed slowly, the magnetization of the movable bit is a continuous, single-valued function of its position. Any work exerted on the movable bit during the first half of the process (until it reaches the center of the transverse field), is compensated by the same amount of

work, but with opposite sign, in the latter half.

In this example, the information stored in a bit is susceptible to thermal fluctuations and tunneling. These phenomena determine a minimum error rate of the copying and a minimum dissipation in each step in the process. However, there is no fundamental theorem that prevents us from making the error probability and dissipation arbitrarily small. Whether one could build this apparatus in the lab was not the main point by Bennett. His examples of reversible measurements are counter-examples to the ones of Gabor and Brillouin. The main point of Bennett was that there is no fundamental law of physics that says that measurements cannot be done without dissipation.

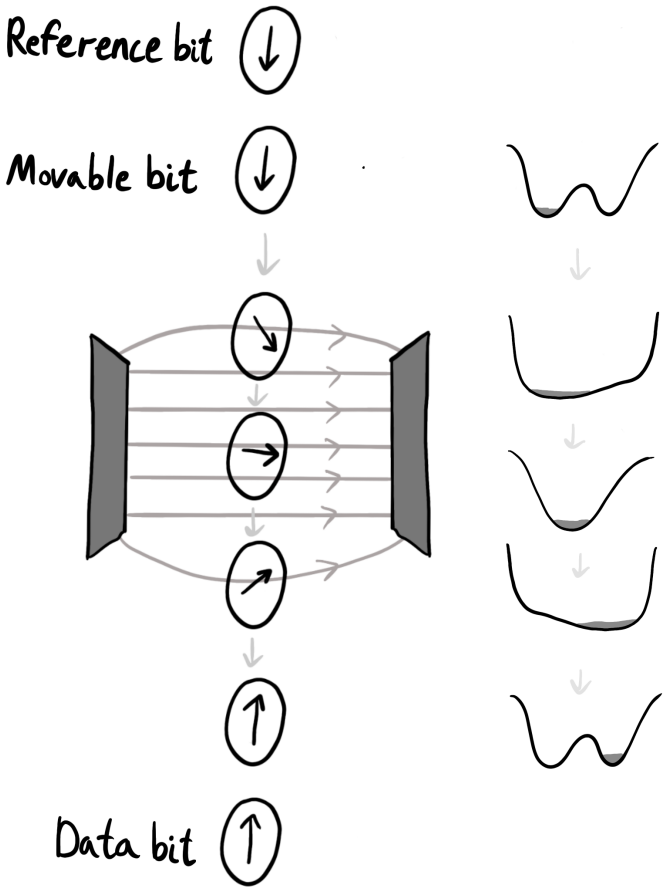

Figure 2.6: Illustration of a reversible measurement using a single-domain ferromagnet. A movable bit, initially in the reference state 0, is brought through a transverse magnetic field, and mapped into the same state as the data bit. The right side shows how the probability density of the movable bit is continually deformed from its initial concentration in the 0 state, until it occupies the 1 state, in agreement with the data bit. Illustration adapted from [17].

2.3.3 Erasure

If measurements can in principle be performed without dissipation, and the difficulty of operating a real Maxwell demon due to fluctuations is not a fundamental prohibition of their existence, how can the second law of thermodynamics be saved? The contemporary view has its origin in Landauer's information erasure principle [26]. He was studying heat generation in computing processes and argued that logically irreversible operations are always associated with physical irreversibility, which requires a minimal heat generation. This idea was applied to the Szilard engine by Bennet [24, 17], who argued that the result of the measurement that the demon performs has to be stored somewhere. Consider a demon with a memory in some initial known standard state S. After measuring the position of the particle in the Szilard engine, and performing the isothermal expansion, we extract an amount of work $k_B T \ln 2$, while reducing the entropy of the heat bath by the same amount. However, the memory of the demon is now in an unknown state, either L (left) or R (right), which has increased its entropy by $k_B \ln 2$. Therefore, when considering the state of the universe, i.e., the combined system of the demon, engine, and environment, the net entropy production is zero. Moreover, the state of the universe is not the same as it was initially, since the state of the demon has changed from a known state S, to an unknown state which is either R or L.

In order to reset the state of the universe to its initial state, so that the engine can operate cyclically, we have to erase the information stored in the demon's memory. This erasure, a two-to-one mapping of the demon's physical state, is a logically irreversible operation, which according to Landauer cannot be accomplished without heat dissipation. Consider the combined cycle of the Szilard engine and the demon's memory, as shown in Fig. 2.7. The left side shows the operation protocol of the Szilard engine, with barrier insertion, measurement and the isothermal barrier expansion. The state of the demon's memory is denoted by S, L or R, and the phase space of the Szilard engine from the point of view of an external observer who does not know the result of the demons measurement, is shown on the right side of the figure. In the phase space illustration, the horizontal axis represents the x-coordinate of the particle, while the vertical axis represents the state of the memory.

For the initial equilibrium state (a), the phase space of the particle occupies

the full box, with an equal probability of finding the particle anywhere in the box. The state of the memory is in its standard state S. After inserting the barrier but before the measurement, the phase space of the particle remains unchanged (except for a small portion in the center, proportional to the width of the barrier, which is infinitesimally small), and the state of the memory is still in its standard state S, as shown in (b). The demon now performs a measurement (c), to find out whether the particle is on the left (L) or right (R) side of the barrier. After the measurement, the demon's state is in either L or R, and the phase space of the particle is concentrated in the corresponding physical states. Based on the information obtained, the demon allows the barrier to expand isothermally (d) while extracting $k_B T \ln 2$ of work. When the expansion is over, the phase space of the particle again fills the whole apparatus (e), and the state of the demon is still in either L or R. The expansion procedure (d-e) depends on which side the particle is found, which is why the demon has to perform a measurement before initiating it. In (e) the physical state of the engine is the same as its initial state, but the information of which side the particle was found is still stored in the demon's memory. In order to put the demon back to its standard state S, this memory has to be deleted, which entails a twofold compression of the demon's phase-space, as shown in (e-f). According to Landauer, this twofold compression is accompanied by an entropy increase of $k_B T \ln 2$, somewhere else in the total system. That is, all the work extracted from heat bath during step (c-e), is converted to heat again when deleting the state of the demon's memory.

In chapter 3.2.2 we go into further detail about the connection between logical and physical irreversibility, as well as where the heat dissipation during erase happens. But for now, we can summarize that the contemporary consensus is that the Szilard engine and Maxwell's demon does not violate the second law of thermodynamics, because one has to take into account the cost of erasing the information the demon obtains from the measurement. For the full cycle of measurement, expansion, and erasure, the minimum entropy production is zero, which corresponds to the lower bound of the second law of thermodynamics, $\Delta S \geq 0$. Landauer and Bennett pointed to a deep connection between information theory and physics. Maxwell's demon has not only been significant due to the challenge it posed to physicists view on the second law of thermodynamics, but also because the resolution of the paradox, and the research surrounding it, revealed the physical implications of information processing in both classical and quantum systems.

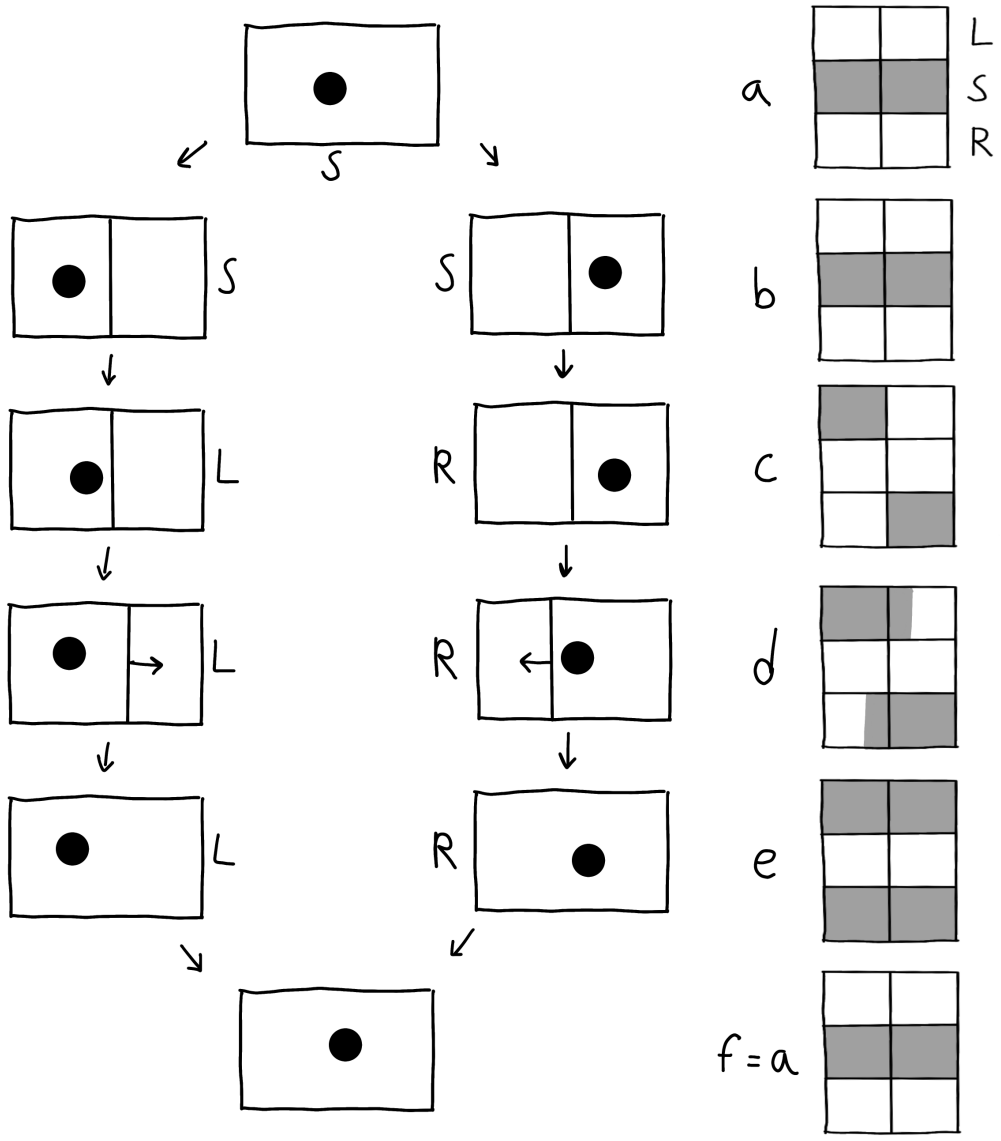

Figure 2.7: Illustration of the combined system of the Szilard engine and demon memory. The left side shows the operation procedure of the Szilard engine, while the right side shows the corresponding phase space evolution of the combined system. Here L and R denotes the outcome of the demons measurement, while S is the initial standard state.

Chapter 3

Statistical mechanics and information theory

In the first section of this chapter, we summarize the most important concepts from statistical mechanics. At the heart of the discussion around Maxwell's demon, lies the concept of information and its physical embodiment; both logical and physical irreversibility, what erasing and obtaining information implies for physical systems, and how these concepts relate to Maxwell's demon-like systems are discussed in the later parts of this chapter. This chapter serves as an introduction to the topics that are needed to understand the rest of this thesis and the associated research articles.

3.1 Statistical mechanics

3.1.1 Foundations of statistical mechanics

In classical mechanics, the time-evolution of a system is described by Hamiltonian dynamics. If we want to describe the behavior of systems with a large number of degrees of freedom, such as an N-particle gas, it is convenient to consider its phase space. The phase space is an imagined space, where each degree of freedom has its own axis. Thus, the phase space of a three-dimensional N-particle gas has 3N axes to specify the coordinates of each particle (x, y, z), and 3N axes to specify the momentum of each particle (p_x, p_y, p_z). A specific point in the 6N-dimensional phase space corresponds to one unique microstate. Under Hamiltonian dynamics, this point moves around in the phase space, as the state of the system changes.

There are not many systems where we have access to the exact microstate. If you are given a container of gas, it would be impossible for you to determine the exact position and momentum of every particle in it. Thus, in the macroscopic world, we deal with macroscopic variables. In general, a macrostate of a system is defined by the properties which we can reliably measure. For a simple ideal gas, this is its temperature T, volume V and pressure p. For magnetic systems, we would include the magnetization M, and for liquids, the surface tension γ. All microstates that correspond to a given macrostate, constitutes a volume in phase space. Within this volume, we can assign a probability distribution to the points in the phase space. The exact distribution we assign depends on what information we have about the system, but the goal is that this probability distribution gives us the probability for the system to be in the corresponding microstate.

If we consider again the N-dimensional gas, a specific point in phase space is specified by 6N independent variables; the N three-dimensional momentum vectors $p_N = (p_1, \ldots, p_N)$, and the N three-dimensional coordinate vectors $q_N = (q_1, \ldots, q_N)$. If the state-vector $x_N = (p_N; q_N)$ is known at one time, it is known for all times, due to deterministic Hamiltonian evolution. Given the Hamiltonian $H_N \equiv H(x_N, t)$, we can find the time evolution of the system using Hamilton's equations,

$$\frac{dp_i}{dt} = -\frac{dH_N}{dq_i} \qquad \text{and} \qquad \frac{dq_i}{dt} = \frac{dH_N}{dp_i}. \tag{3.1}$$

The state-vector x_N traces out a trajectory in phase space, as it evolves in time. Since Hamiltonian dynamics defines a unique past and future for a given state x_N, it follows that the trajectory can not cross itself. If it could, then Hamiltonian evolution would be indeterministic. If we lack complete knowledge of the system, we have to consider x_N to be a stochastic variable, and associate a probability density, $\rho(x_N, t)$ to the phase space. The probability that the state is found in a volume element dx_N around x_N at time t, is then given by $\rho(x_N, t)dx_N$. Since the state must always lie somewhere in the phase space, the probability density has to be properly normalized:

$$\int_\Gamma \rho(x_N, t)dx_N = 1, \tag{3.2}$$

where $\int_\Gamma$ indicates integration over the full phase space. The probability to find

the system in a region R is then given by

$$P(x_N \in R) = \int_R \rho(x_N, t) dx_N. \tag{3.3}$$

We can view this probability density in phase space as an incompressible fluid, that flows according to Hamiltonian dynamics. Therefore we can use fluid mechanics to find its equation of motion; the Liouville equation.

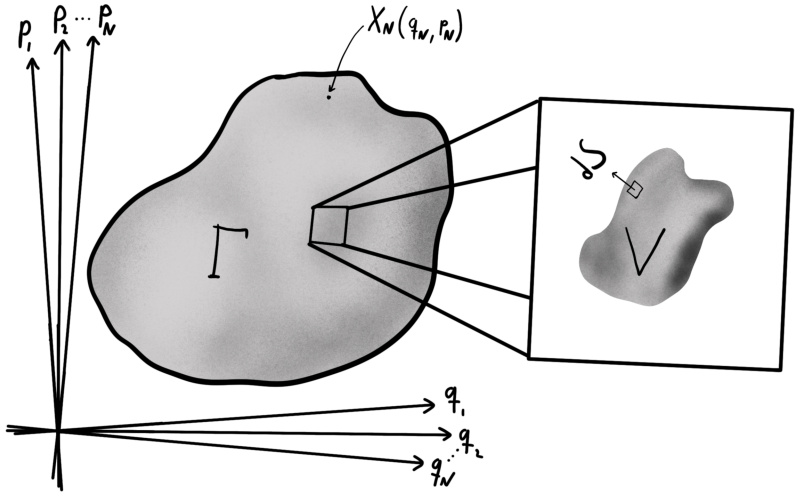

Figure 3.1: Representation of a 2Nd dimensional phase space, $\rho(x_N, t)$, where d is the spacial dimension of the system and N is the number of particles. The total phase space we consider is given by Γ, while a small volume element of that is V. The differential area-element normal to the surface of V is given by dS.

Consider a small volume element V with surface area S,at a fixed point in phase space, as shown in Fig. 3.1. The total probability is conserved, so any change in the probability to find the state in this volume,

$$\frac{d}{dt} P(x_N \in V) = \frac{\partial}{\partial t} \int_V \rho(x_N, t) dx_N, \tag{3.4}$$

is also given by the flow of probability through it,

$$\frac{d}{dt} P(x_N \in V) = - \oint_S \rho(x_N, t) \dot{x}_N \cdot dS. \tag{3.5}$$

Here $\dot{x}_N$ is the velocity of the state-vector, and dS is the area-element normal to the surface S. We can now use Gauss's theorem, which transforms the surface integral to a volume integral, to obtain

$$\frac{\partial}{\partial t}\int_V \rho(x_N,t)dx_N = -\int_V \boldsymbol{\nabla}_{x_N}\cdot\left[\rho(x_N,t)\,\dot{x}_N\right]dx_N, \tag{3.6}$$

where $\boldsymbol{\nabla}_{x_N} = (\partial_{q_1},\ldots\partial_{q_N},\partial_{p_1},\ldots,\partial_{p_N})$ is the gradient with respect to all the phase space variables. Since the volume area V is independent of time, we can take the time-derivative inside the integral on the left side. The arguments of the integral therefore have to be the same, giving us

$$\frac{\partial}{\partial t}\rho(x_N,t) + \boldsymbol{\nabla}_{x_N}\cdot[\rho(x_N,t)\dot{x}_N] = 0. \tag{3.7}$$

We can calculate the divergence term to get

$$\boldsymbol{\nabla}_{x_N}\cdot[\rho(x_N,t)\dot{x}_N] = \dot{x}_N\cdot\boldsymbol{\nabla}_{x_N}\rho(x_N,t) + \rho(x_N,t)\boldsymbol{\nabla}_{x_N}\cdot\dot{x}_N,$$

and if we now use Hamilton's equations (Eq. 3.1), we see that

$$\boldsymbol{\nabla}_{x_N}\cdot\dot{x}_N = \sum_{i=1}^{N}\left(\frac{\partial\dot{q}_i}{\partial q_i}+\frac{\partial\dot{p}_i}{\partial p_i}\right) = \sum_{i=1}^{N}\left(\frac{\partial^2 H_N}{\partial q_i\partial p_i}-\frac{\partial^2 H_N}{\partial p_i\partial q_i}\right) = 0. \tag{3.8}$$

From Eq. (3.7) we therefore get

$$\frac{\partial}{\partial t}\rho(x_N,t) + \dot{x}_N\cdot\boldsymbol{\nabla}_{x_N}\rho(x_N,t) = 0 \tag{3.9}$$

Since the total time derivative is defined as

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \dot{x}_N\boldsymbol{\nabla}_{x_N}, \tag{3.10}$$

we see that if we pick a specific point in phase space x_N, and follow its trajectory as the phase space evolves it time, the probability density in the neighborhood of that point remains constant:

$$\frac{d}{dt}\rho(x_N,t) = 0. \tag{3.11}$$

We can further rewrite Eq. (3.9) into a more familiar form by using Hamilton's

equation.

$$\begin{aligned}\frac{\partial}{\partial t}\rho(x_N,t) &= -\dot{x}_N\cdot\boldsymbol{\nabla}_{x_N}\rho(x_N,t)\\ &= -\sum_{i=1}^{N}\left(\frac{\partial\dot{q}_i}{\partial t}\frac{\partial}{\partial q_i}+\frac{\partial\dot{p}_i}{\partial t}\frac{\partial}{\partial p_i}\right)\rho(x_N,t)\\ &= -\sum_{i=1}^{N}\left(\frac{\partial H_N}{\partial p_i}\frac{\partial}{\partial q_i}-\frac{\partial H_N}{\partial q_i}\frac{\partial}{\partial p_i}\right)\rho(x_N,t). \qquad (3.12)\end{aligned}$$

This equation is known as Liouville's equation, and it is often written using the Poisson bracket notation:

$$\frac{\partial}{\partial t}\rho(x_N,t) = -\left\{\rho(x_N,t),H_N\right\}. \qquad (3.13)$$

It is the equation of motion for the probability density in phase space, and from it we can solve any dynamical Hamiltonian system, given that we know the initial probability density $\rho(x_N,0)$. A probability density that does not depend on time, $\partial_t\rho(x_N,t)=0$, is associated with a system at equilibrium. The condition that makes both Liouville's equation and the stationary probability density compatible is clearly

$$\{\rho(x_N,t),H_N\}=0. \qquad (3.14)$$

Once choice of $\rho(x_N,t)$ that satisfies this equation is one that does not depend on x_N. In other words

$$\rho(x_N,t)=\text{const.} \qquad (3.15)$$

In general, the Hamiltonian flow of the phase space density makes an initially smooth phase space density quickly evolve into an extremely complicated structure, with tendrils going in all directions in phase space. An illustration of this is shown in Fig. 3.2, where an initially spherical phase space density evolves into a complicated structure. However, no matter how complicated the structure becomes, its total volume remains the same.

3.1.2 Ensemble theory

Ensemble theory is the foundation that all of statistical mechanics is built upon. The probability density $\rho(x_N,t)$ can be interpreted as an ensemble of microstates belonging to the same macrostate. If we imagine we have M identical copies of

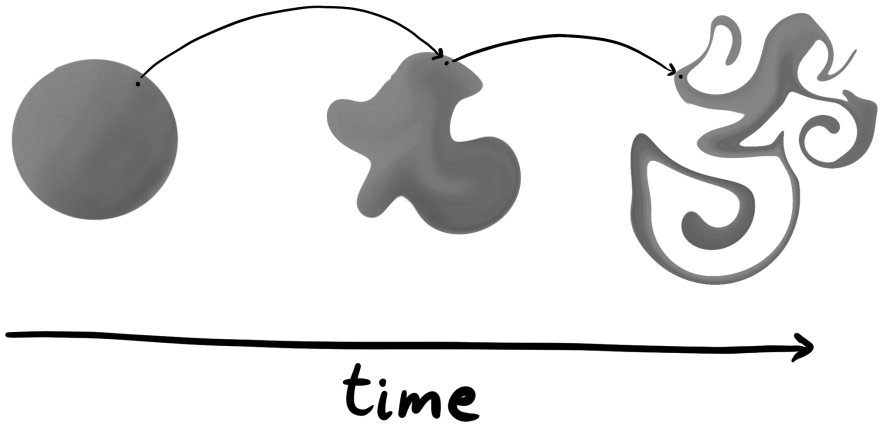

Figure 3.2: Illustration of the Hamiltonian flow of a phase space density according to Liouville's theorem.

a three-dimensional ideal gas (6N dimensional phase space), each member of the ensemble is a vector pointing to a point in the phase space. The density of these representative points in phase space is then given by $M\rho(x_N, t)$.

The ensemble average of a function $f(x_N)$ is defined as

$$\langle f \rangle_\Gamma = \frac{\int_\Gamma f(x_N)\, \rho(x_N, t)\, dx_N}{\int_\Gamma \rho(x_N, t)\, dx_N}, \tag{3.16}$$

where the integration extends over the full phase space Γ. In general f can be an explicit function of time $f = f(x_N, t)$, which makes the ensemble average time-dependent as well. The ensemble is stationary if

$$\frac{\partial}{\partial t}\rho(x_N, t) = 0, \tag{3.17}$$

and for such an ensemble the average value of any function f will be time independent. Stationary ensembles correspond to equilibrium distributions, and the condition that ensures that a system is both in equilibrium and obeys Hamiltonian dynamics can be found by combining Eq. (3.17) and Eq. (3.13) to obtain

$$\{\rho(x_N, t), H_N\} = 0. \tag{3.18}$$

A final thing we need in order to define the different ensembles, is the concept of ergodicity. We can define the time average of a function $f(x_N)$ as

$$\langle f \rangle_T = \lim_{T\to\infty} \frac{1}{T} \int_{t_0}^{t_0+T} f(x_N)\, dt. \tag{3.19}$$

The ergodic hypothesis states that for a given equilibrium macro state, the time spent by the system in some region of phase space is proportional to the volume of the region. This implies that all corresponding micro states are equally probable of a long period of time, which makes the ensemble average equal to the time average

$$\langle f \rangle_\Gamma = \langle f \rangle_T . \tag{3.20}$$

The exact time scale where the ergodic hypothesis becomes valid depends on the macroscopic system in question. For some system the time it takes to explore the full phase space can be so large that the equilibrium state exhibit ergodicity breaking. We also see that the probability to find a macro state in some specific region of phase space, is proportional to the area of the region.

The microcanonical ensemble

The simplest ergodic stationary state, is given by a Hamiltonian of constant energy $H(x_N) = E$. This equation defines a hypersurface in the phase space. For a gas with 6N dimensional phase space, the energy hypersurface spans 6N-1 dimension. From theorem Eq. (3.14), we see that the probability density compatible with this Hamiltonian is one that is constant everywhere on the hypersurface. And from the ergodic theorem, we know that the probability to find the system in a region R in phase space, is proportional to the area of that region, which we can normalize using the total area of the hypersurface

$$P(x_N \in R) = \frac{\int_R \delta\left(H(x_N) - E\right) dx_N}{\int_\Gamma \delta\left(H(x_N) - E\right) dx_N} = \frac{\Omega(R)}{\Omega(E)}. \tag{3.21}$$

Here $\Omega(R)$ is the area of the region R, while $\Omega(E)$ is the area of the full energy hypersurface. We can then write down the normalized probability distribution of

the energy surface as

$$\rho(x_N, E) = \begin{cases} \frac{1}{\Omega(E)}, & \text{for } H(x_N) = E \\ 0, & \text{otherwise.} \end{cases} \tag{3.22}$$

This probability distribution constitutes the microcanonical ensemble, and represents a closed system with constant energy, where we equal a priori probabilities for the possible micro states.

The canonical ensemble

Most thermodynamic systems do not have an exactly fixed energy. Even a closed system at equilibrium will exchange heat with its environment, in such a way that the energy of the system fluctuates around a mean value. To find the equilibrium distribution of such and ensemble we maximize the Gibbs entropy,

$$S = -k_B \int dx_N \; \rho(x_N) \; \log \rho(x_N). \tag{3.23}$$

This is identical to the Shannon entropy (derived in chapter 3.2.1), with $K = k_B$. Since the maximization is constrained by the normalized probability and the average energy,

$$\int_\Gamma dx_N \rho(x_N) = 1, \qquad \int_\Gamma dx_N \rho(x_N) H_N = \langle E \rangle \,, \tag{3.24}$$

we use the method of Lagrange multipliers to obtain

$$\lambda_1 - k_B + \lambda_2 H_N - k_B \log \rho(x_N) = 0, \tag{3.25}$$

where $\lambda_{1/2}$ are the Lagrange multipliers. This gives us

$$\rho(x_N) = \exp\left(\frac{\lambda_1}{k_B} - 1 + \frac{\lambda_2}{k_B} H_N\right). \tag{3.26}$$

To determine the Lagrange multipliers we first use the normalized probability condition and obtain

$$\int_\Gamma \exp\left(\frac{\lambda_2}{k_B} H_N\right) = \exp\left(1 - \frac{\lambda_1}{k_B}\right). \tag{3.27}$$

Next, we take Eq. (3.25), multiply it by $\rho(x_N)$ and integrate over Γ. This leaves us with

$$-k_B \int_\Gamma dx_N e^{\left(\frac{\lambda_2}{k_B} H_N\right)} + \lambda_2 \langle E \rangle + S = 0. \tag{3.28}$$

Comparing this equation to the definition of the Helmholtz free energy $F - U + TS = 0$, we see that $\lambda_2 = -1/T = -\beta$. Putting it all back into Eq. (3.26) we finally obtain the probability density for the canonical ensemble;

$$\rho(x_N) = \frac{e^{-\beta H_N}}{\int_\Gamma dx_N \; e^{-\beta H_N}} = \frac{e^{-\beta H_N}}{Z_\Gamma}. \tag{3.29}$$

The function Z_Γ is the canonical partition function, and can be considered a normalization constant for the probability density $\rho(x_N)$.

The grand canonical ensemble

The grand canonical ensemble is derived in an almost identical way as above, only now we maximize the entropy with an additional constraint, on the average number of particles $\int N\rho(x_N)\, dx_N = \langle N \rangle$. The probability density in the grand canonical ensemble becomes

$$\rho(x_N) = \frac{e^{-\beta(H_N - \mu N)}}{\int_\Gamma dx_N \; e^{-\beta(H_N - \mu N)}}, \tag{3.30}$$

where μ is the chemical potential.

3.2 Information and entropy

3.2.1 Shannon entropy

Claude Shannon, while working at Bell Telephone Laboratories, developed in 1948 a mathematical measure of uncertainty, to quantify the loss of information in phone-line signals [27]. Supposedly while working on this measure he visited Von Neumann, and they had the following discussion:

> *My greatest concern was what to call it. I thought of calling it information, but the word was overly used, so I decided to call it uncertainty. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, "You should call it entropy, for two*

> *reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage".*

Shannon followed Von Neumann's advice, and called his measure the Shannon entropy. E.T. Jaynes has a clear derivation of Shannon entropy that we will follow from now on [28]. Assume we have a variable x that can take on discrete values $(x_1 \dots x_n)$. The process that determines what value x assumes can be represented by the corresponding probabilities $(p_1 \dots p_n)$, where p_i represents the probability that $x = x_i$. The goal is to derive a quantity $H(p_1 \dots p_n)$, which uniquely measures the amount of uncertainty represented by this probability distribution. Or in other words, a function that quantifies our lack of information about a system. It might seem difficult to create an unique and consistent measure of uncertainty. Remarkably, only by using three elemental conditions of consistency we can show that this quantity H is what we now call Shannon entropy. The three conditions are:

(1) H has to be a continuous function of the p_i's, or else an arbitrarily small change in their value would lead to a large change in the amount of uncertainty.

(2) If all p_i are equal, the quantity $h(n) = H(\frac{1}{n} \cdots \frac{1}{n})$ is a monotonic increasing function of n: If you don't know anything about the distribution, your uncertainty can only increase if the number of possible choices increases.

(3) The measure H has to be consistent, meaning that if there is more than one way of calculating its value they all have to give the same answer.

In the opening statement we said that x can assume any of the discrete values $(x_1 \dots x_n)$, thus we can not assign $p_i = 0$ for any x_i. Unless we *know* what value x is e.g., $p_k = 1$, we have to give a finite value for all p_i. But if we know that $p_k = 1$ then we have complete information about the distribution, and a function describing our lack of knowledge is nonsensical.

According to condition (3), we have a choice between giving the probabilities of the events $(x_1 \dots x_n)$ directly, or partitioning them in groups. We can group the first k of them, such that the group probability is $\omega_1 = (p_1 + \cdots + p_k)$, then group the next m so that the probability is $\omega_2 = (p_{k+1} + \cdots + p_{k+m})$, and so on. The

amount of uncertainty of the composite events is then $H(\omega_1, \ldots, \omega_N)$, where N is the total number of groups. The conditional probabilities of the events $(x_1 \ldots x_k)$, given the composite event ω_1 is then $(p_1/\omega_1, \ldots, p_k/\omega_1)$. Doing this for all the composite events, eventually brings us to the same state of knowledge as if all the p_i's had been given directly.

$$\begin{aligned} H(p_1 \ldots p_n) &= H(\omega_1 \ldots \omega_r) + \omega_1 H(p_1/\omega_1 \ldots p_k/\omega_1) \\ &+ \omega_2 H(p_{k+1}/\omega_2 \ldots p_{k+m}/\omega_2) + \cdots . \end{aligned} \tag{3.31}$$

That is, the uncertainty given by the p_i's, is the same as the uncertainty of composite events plus the conditional probability of each composite event. As an example, lets say we have $(p_1, p_2, p_3) = (1/2, 1/3, 1/6)$ and decide to form the two following groups; $\omega_1 = p_1 = 1/2$, and $\omega_2 = p_2 + p_3 = 1/2$. We then get

$$\begin{aligned} H(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) &= H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} H\left(\frac{1}{2}\right) + \frac{1}{2} H\left(\frac{1/3}{1/2}, \frac{1/6}{1/2}\right) \\ &= H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} H\left(\frac{2}{3}, \frac{1}{3}\right) \end{aligned} \tag{3.32}$$

Since H is continuous according to condition (1), it is sufficient to determine H for all rational values

$$p_i = n_i / \sum_i n_i, \qquad n_i = \text{integers}. \tag{3.33}$$

We can then regard each probability p_i's as a grouping of n_i equally likely events. We can group together any number of equally likely events, to create a composite event of arbitrary probability. Take as an example $N = 9$ equally likely events, and then form the following $n = 3$ groups; one group of $n_1 = 4$, one group of $n_2 = 3$, and one group of $n_3 = 2$. The composition law, Eq. (3.31) then becomes

$$h(9) = H\left(\frac{4}{9}, \frac{3}{9}, \frac{2}{9}\right) + \frac{4}{9} h(4) + \frac{3}{9} h(3) + \frac{2}{9} h(2), \tag{3.34}$$

where $h(n)$ is shorthand for

$$h(n) = H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right). \tag{3.35}$$

The general form of Eq. (3.31) with this notation becomes

$$h(\sum_{i}^{n} n_i) = H(p_i, \ldots, p_n) + \sum_i p_i h(n_i). \tag{3.36}$$

If we now choose all $n_i = m$, the equation further simplifies to

$$h(mn) = h(m) + h(n), \tag{3.37}$$

which can be shown [27] to have the unique solution

$$h(n) = K \log(n), \tag{3.38}$$

where K is an arbitrary constant. Combining this with Eq. (3.36) we get

$$\begin{aligned} H(p_1, \cdots, p_n) &= K \ln\Big(\sum n_i\Big) - K \sum_i p_i \ln(n_i) \\ &= K \ln\Big(\sum n_i\Big) - K \sum_i p_i \ln\left(p_i \sum_i^n n_i\right) \\ &= K \ln\Big(\sum n_i\Big) - K \sum_i p_i \ln p_i - K \sum_i p_i \ln\left(\sum_i n_i\right) \\ &= -K \sum_i p_i \ln p_i, \end{aligned} \tag{3.39}$$

which is the familiar form of the Shannon entropy, and this is only equation that satisfies the conditions we imposed. It then follows that for a given a probability distribution $(p_1, \cdots, p_n)$, the values of the p_i's that maximizes the Shannon entropy is the least biased and most "honest" description of a system, subject to the constraints imposed by our available information.

We can find the maximum of H, given that the probability is normalized, by using the method of Lagrange multipliers.

$$\begin{gathered} \nabla \left[H(p_1 \ldots p_n) - \lambda G(p_1 \ldots p_n)\right] = 0 \\ \Downarrow \\ \max \left\{H(p_1 \ldots p_n) \mid G(p_1 \ldots p_n) = 0\right\}, \end{gathered} \tag{3.40}$$

where $G(p_1 \ldots p_n) = \sum_i p_i - 1$. Performing the calculation of the gradient along

one dimension p_k, we obtain

$$-\ln p_k - 1 - \lambda = 0 \tag{3.41}$$

$$p_k = e^{-(1+\lambda)} \tag{3.42}$$

which has to apply for all p_k. Putting this into the normalization constraint gives us

$$\sum_i^N e^{-(1+\lambda)} = N e^{-(1+\lambda)} = 1 \quad \rightarrow \quad \lambda = \ln(N) - 1, \tag{3.43}$$

with the final result

$$p_k = e^{-(1+\ln N - 1)} = e^{-\ln N} = \frac{1}{N}. \tag{3.44}$$

The implication is that $p_i = \frac{1}{N}$ is the *least biased* probability distribution for the points $(p_1 \ldots p_n)$, and it indicates that we don't know anything about the distribution other than how many possible outcomes there are.

3.2.2 Thermodynamic and logical reversibility

Consider a general system where the total phase space is Γ, and the phase space coordinates are described by the vectors $\gamma \in \Gamma$. The system is surrounded by a heat bath at inverse temperature $\beta = 1/k_B T$. A transformation that maps some initial phase space distribution $\{ \gamma_i \} \equiv \Gamma_i \subset \Gamma$ to some final distribution $\{ \gamma_f \} \equiv \Gamma_f \subset \Gamma$, is then due to some physical process. The Shannon entropy of the initial and final state is given by

$$S_i = -\int_{\gamma \in \Gamma_i} d\gamma \; p(\gamma) \ln p(\gamma) \quad \text{and} \quad S_f = -\int_{\gamma \in \Gamma_f} d\gamma \; p(\gamma) \ln p(\gamma), \tag{3.45}$$

where $p(\gamma)$ is the probability of the state represented by the phase space point γ. If Q is the average heat absorbed by the system under the transformation, the total entropy production (i.e., system + environment) is then given by

$$\Delta S_{tot} = \underbrace{(S_f - S_i)}_{\Delta S} - \beta Q. \tag{3.46}$$

According to the second law of thermodynamic, total entropy change is bounded below at zero

$$\Delta S_{tot} \geq 0 \quad \rightarrow \quad \Delta S \geq \beta Q. \tag{3.47}$$

A physical process which achieves equality in this bound, is considered a thermodynamically reversible process. Notice that the flow of entropy between system and bath is possible for reversible processes, if the amount of heat absorbed by the system is equal to its entropy change. This is because the absorption of heat by the system results in a decrease in the environment entropy according to $\Delta S_{env} = -Q\beta$.

If the phase space points of our system is distributed according to the canonical distribution, the probabilities $p(\gamma)$ is given by

$$p(\gamma) = e^{\beta\left(F - E(\gamma)\right)}, \tag{3.48}$$

where $E(\gamma)$ is the energy associated with the phase space point γ, and $F = -\ln Z$ is the free energy associated with the distribution of phase space points $\{\,\gamma\,\}$. With this probability the entropy of the initial and final state becomes

$$S_{(i/f)} = \beta\left(U_{(i/f)} - F_{(i/f)}\right), \tag{3.49}$$

where $U_{(i/f)} = \langle E_{(i/f)} \rangle_C$ is the canonical ensemble average of the energy. Using the first law of thermodynamics, $\Delta U = \Delta W + \Delta Q$, where W is the average work performed on the system, we find that the second law of thermodynamic in this form becomes

$$\Delta S_{tot} = \beta\left(W - \Delta F\right) \geq 0 \quad \rightarrow \quad W \geq \Delta F, \tag{3.50}$$

where $\Delta F = F_f - F_i$. We see that if the input work we perform on the system is equal to its change in its free energy, the process is reversible.

Phase space trajectories can not cross each other, because if they could the phase space point at the intersection does not have a deterministic Hamiltonian evolution. The point could evolve according to either trajectory, so we would lose information about its past. This concept is closely related to logical reversibility. Consider a set of logical input states I, and logical output states O. Lets for simplicity consider one single bit of information, that can be in one of two logical

states $\{0, 1\}$. A logical process, or a computation C, can then be described as a transformation between the input state and the output state $C : I \to O$. An example of an irreversible process is then the ERASE operation, which is defined by

$$\text{ERASE}: \quad 0 \to 0, \quad 1 \to 0. \tag{3.51}$$

No matter which state you were in (0 or 1), you end up in the same state (0), and lose any information about the past. An example of a reversible process is the NOT operation, which is defined as

$$\text{NOT}: \quad 0 \to 1, \quad 1 \to 0. \tag{3.52}$$

In this case, given the output, you always know the input. A logically reversible process can be defined as one that, for any output logical state, a unique input logical state exists [29]. Meaning that for every logical state in O, there exists a reversal of C, which is defined as $C^{-1} : O \to I$.

Now let the input and output states be two probability distributions instead of a single bit. We defined them as $P_O(n_o)$ for $n_o \in O$ and $P_I(n_i)$ for $n_i \in I$, with normalized probabilities. After the operation C, the distribution on O is given by

$$P_0(n_o) = \sum_{n_i:\ C(n_i)=n_o} P_I(n_i), \tag{3.53}$$

where the sum is taken over all n_i which satisfies $C(n_i) = n_o$. If the process is reversible, then there is one unique n_i for each n_o, giving us

$$P_0(n_o) = P_I(C^{-1}(n_o)). \tag{3.54}$$

The input and output logical entropies are given by

$$H_I = -\sum_{n_i \in I} P_I(n_i) \ln P_I(n_i), \tag{3.55}$$

and

$$H_O = -\sum_{n_o \in O} P_O(n_o) \ln P_O(n_o). \tag{3.56}$$

For reversible operations, defined by Eq. (3.54), we see that the logical entropy

does not change

$$\begin{aligned} H_O &= -\sum_{n_o \in C(O)} P_I(C^{-1}(n_o)) \ln P_I(C^{-1}(n_o)) \\ &= -\sum_{n_i \in I} P_I(n_i) \ln P_I(n_i) = H_I. \end{aligned} \tag{3.57}$$

In the general case, including non-reversible operation, the entropy difference becomes

$$H_O - H_I = \sum_{n_o} P_O(n_0) \sum_{n_i:\ C(n_i)=n_o} \frac{P_I(n_i)}{P_O(n_o)} \ln \frac{P_I(n_i)}{P_O(n_o)} \tag{3.58}$$

3.3 Erasing information: Landauer's principle

As discussed in the introduction, Landauer's solution to the apparent violation of the second law of thermodynamics by the Szilard engine was the fact that one has to erase the information obtained by the measurement [26, 30, 31]. All physical systems designed to perform logical operations have specific physical states (microstates) which correspond to the logical states. A one-bit memory can be modeled as a single-particle-box with a barrier in the center, as shown in Fig. 3.3(a). The two logical states are a particle found on the left side of the barrier (0) or a particle found on the right side of the barrier (1). In this model the logical states $\{\, 0, 1 \,\}$ correspond to the physical states

$$0 \equiv \{\, x \in [-L/2, 0], |p| = \sqrt{2mE} \,\}, \tag{3.59}$$

and

$$1 \equiv \{\, x \in [0, L/2], |p| = \sqrt{2mE} \,\}. \tag{3.60}$$

Landauer argued that logically irreversible processes, which reduce the logical state space, must therefore also compress the physical state space. This compression of phase space results in an increase in entropy, in the form of heat dissipation [24, 25]. An example of a logical irreversible process is the ERASE operation discussed earlier ($0 \to 0, 1 \to 0$). The physical implementation of this protocol on the SPB memory is shown in Fig. 3.3(b). The memory is initially in either of the two logical states $\{\, 0, 1 \,\}$. We then remove the barrier from the center of the box, and insert it in the far right-hand side of the box. While the barrier back towards the center, the collisions between the particle and the barrier exerts an

effective pressure on the barrier. Therefore an amount of work is required to push the barrier, which is transferred to the heat bath via the thermal contact between the particle and environment. When the barrier reaches the center of the box, the particle is always found in a physical state corresponding to the logical state 0.

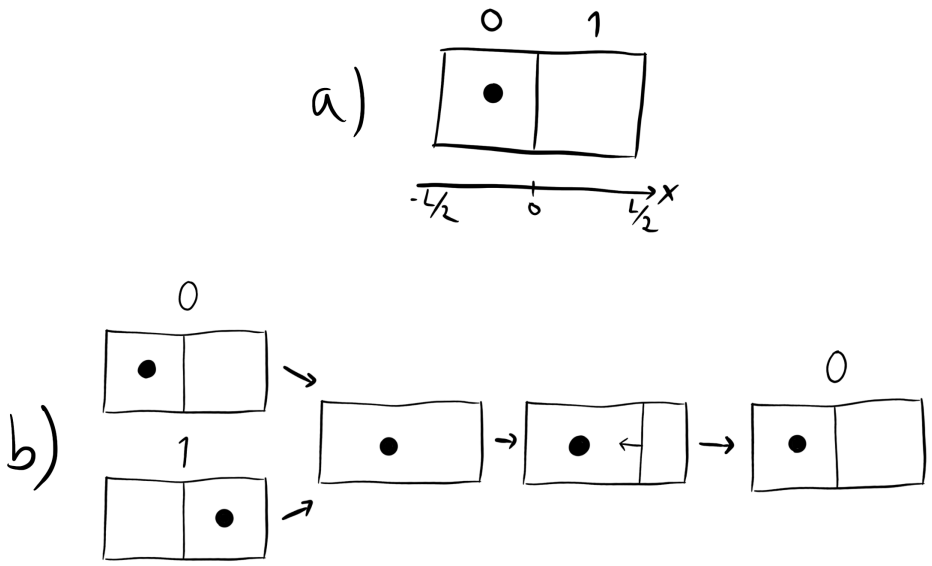

Figure 3.3: Illustration of a binary memory, modeled as a SPB of width L, and two logical states; left side of the barrier (0) and right side of the barrier (1). In b) we show a physical implementation of the ERASE operation.

Before the erasure, the probability of 0 and 1 are equally 1/2, giving a logical entropy $H_i = \ln 2$. After the erasure has been performed, the probability of 0 is 1, so the logical entropy is $H_f = 0$. The difference in logical entropy is therefore $\Delta H = H_f - H_i = -\ln 2$. Since the logical entropy has to be treated on the same level as physical entropy, we have $\Delta S = \Delta H$, and from the second law of thermodynamics (Eq. (3.47)) we obtain

$$-\ln 2 \geq \beta Q \tag{3.61}$$

where Q is the heat dissipated into the environment. Since the internal energy does not change during the isothermal erasure we have, according to the first law of thermodynamics, $W = -Q$. Therefore the work needed to erase one bit of

information is given by

$$W \geq k_B T \ln 2. \tag{3.62}$$

This equation is known as Landauer's principle. Equality is achieved if the erasure is performed adiabatically, in such a way that the memory is always in equilibrium with the environment while we push the barrier towards the center. A quasi-static isothermal compression requires an amount of work given by

$$W = \int_{V/2}^{V} \frac{k_B T}{V'} dV' = k_B T \ln 2, \tag{3.63}$$

and is , therefore, an example of a physical erasure protocol that reaches equality in the Landauer bound. The Landauer principle has in recent years been experimentally verified in a number of different systems [32, 33, 34].

In general, a logical state does not have a one-to-one mapping to a unique physical state. Rather, a logical state is a subset of the full phase space, $\Gamma_{0/1} \subset \Gamma$, and corresponds to many different microstates. By definition $\Gamma = \Gamma_0 \cup \Gamma_1$, and $\Gamma_0 \cap \Gamma_1 \equiv \emptyset$. If this was not the case, the two logical states would have indeterminate members which could not be definitely associated with either state. In the previous case the logical state 0 is associated with the subspace $\Gamma_0 : \{ x \in [-L/2, 0] \}$, while the logical state 1 is given by $\Gamma_1 : \{ x \in [0, L/2] \}$. Ignoring the irrelevant y-coordinate and momentum $\vec{p} = p_x\vec{x} + p_y\vec{y}$, we denote the probability distribution of the total phase space by $P(x)$. The probability distribution of the logical states, P_L, is then given by

$$P_L(i) = \int_{x \in \Gamma_i} P(x)\, dx, \qquad i = 0, 1. \tag{3.64}$$

The conditional probability of the microstate x given the logical state i is therefore

$$P(x|i) = P(x)/P_L(i). \tag{3.65}$$

The total entropy, S, is given by the integral over the total phase space

$$S = \int_{\Gamma} P(x) \ln P(x)\, dx, \tag{3.66}$$

while the logical entropy is given by

$$H = -\sum_i P_L(i) \ln P_L(i). \tag{3.67}$$

Following the discussion in Section 3.2.1, we can group the microstates into composite events, i.e. the logical states $\Gamma_{0/1}$ in this case. The total entropy can then be written as the entropy of the logical states, plus the conditional entropy $S(\Gamma_i|i)$ weighted by the logical state probabilities

$$S = -\sum_i P_L(i) \ln P_L(i) - \sum_i P_L(i) S(\Gamma_i|i), \tag{3.68}$$

where

$$S(\Gamma_i|i) = \int_{x\in\Gamma_i} P(x|i) \ln P(x|i)\, dx. \tag{3.69}$$

We see that the total entropy can be decomposed into two terms, where one is the logical entropy H, and the other is the average conditional entropy $S_{in} = \sum_i P_L(i)\, S(\Gamma_i|i)$, which we identify as the internal physical entropy in the logical subspaces

$$S = H + S_{in}. \tag{3.70}$$

Using this decomposed version of the total entropy, we can calculate contribution of each term for an ERASURE operation. For this operation the initial logical probability distribution is $P_L(0) = P_L(1) = 1/2$, while the final one is $P'_L(0) = 1$ and $P'_L(1) = 0$, which gives us a change in logical entropy $\Delta H = -\ln 2$. The change in internal entropy is

$$\begin{aligned}
\Delta S_{in} &= -\sum_i P'_L(i) S'(\Gamma_i|i) + \sum_i P_L(i) S(\Gamma_i|i) \qquad (3.71)\\
&= -S'(\Gamma_0|0) + \frac{1}{2} S(\Gamma_0|0) + \frac{1}{2} S(\Gamma_1|1)\\
&= -\int_{x\in\Gamma_0} \frac{P'(x)}{P'_L(0)} \ln \frac{P'(x)}{P'_L(0)} + \frac{1}{2}\int_{x\in\Gamma_0} \frac{P(x)}{P_L(0)} \ln \frac{P(x)}{P_L(0)} + \frac{1}{2}\int_{x\in\Gamma_1} \frac{P(x)}{P_L(1)} \ln \frac{P(x)}{P_L(1)}\\
&= -\int_{x\in\Gamma_0} dx\; P'(x) \ln P'(x) + \int_{x\in\Gamma} dx\; P(x) \ln P(x) + \int_{x\in\Gamma} dx\; P(x) \ln(2).
\end{aligned}$$

If we assume the initial and final phase space probabilities are equilibrium distri-

butions, with $P(x) = \frac{1}{L}$ and $P'(x) = \frac{1}{L/2}$ we obtain

$$\begin{aligned} \Delta S_{in} &= -\int_{-L/2}^{0} dx\, \frac{2}{L} \ln \frac{2}{L} + \int_{-L/2}^{L/2} dx\, \frac{1}{L} \ln \frac{1}{L} + \ln 2 \\ &= -\ln \frac{2}{L} + \ln \frac{1}{L} + \ln 2 = 0 \end{aligned} \tag{3.72}$$

Therefore the total change in entropy when adiabatically erasing one bit of information is

$$\Delta S = \Delta H + \Delta S_{in} = -\ln 2, \tag{3.73}$$

and the generalized Landauer principle can be expressed as

$$\Delta H + \Delta S_{in} \geq \beta Q. \tag{3.74}$$

3.4 Obtaining information: Measurement

A measurement is to make a copy of the state of a system onto a memory. For the measurement of the state of a Szilard engine, we need a binary memory. We consider the total phase space (system + memory) to be $\Gamma = \Gamma_S \cup \Gamma_M$, where Γ_S and Γ_M is the phase space of the system and memory, respectively. Let $s \in S = \{\, 0, 1 \,\}$ and $m \in M = \{\, 0, 1 \,\}$ be the logical states of the system and memory, respectively. Their physical states is denoted by $x_s \in \Gamma_S$ and $x_m \in \Gamma_M$. The conditional probability of finding the total system in the physical state (x_s, x_m) given the logical states (s, m) is then $P(x_s, x_m|s, m)$, and the probability of the physical state is given by

$$P(x_s, x_m) = \sum_{s,m} P(x_s, x_m|s, m) P(s, m) \tag{3.75}$$

To characterize the correlation between the memory and the system, we introduce the mutual information. The mutual information quantifies how much information we obtain about one subsystem when observing another subsystem; If the mutual information is zero, the state of the memory and system is independent of each other. The mutual information between the physical states are given by

$$I_{in}(\Gamma_S; \Gamma_M) = S_{in}(\Gamma_S) + S_{in}(\Gamma_M) - S_{in}(\Gamma), \tag{3.76}$$

while for the logical states we have

$$I_H(S;M) = H(S) + H(M) - H(S \otimes M), \tag{3.77}$$

where $S \otimes M$ is the total logical state, i.e. $00, 01, 10, 11$. The mutual information between the internal states, given the logical states s and m, are given by

$$I_{in}(\Gamma_S;\Gamma_M|s,m) = S_{in}(\Gamma_S|s) + S_{in}(\Gamma_M|m) - S_{in}(\Gamma|s,m). \tag{3.78}$$

Taking the average over s and m, we obtain

$$I_{in}(\Gamma_S;\Gamma_M|L) = S_{in}(\Gamma_S|S) + S_{in}(\Gamma_M|M) - S_{in}(\Gamma|S \otimes M). \tag{3.79}$$

In a similar way that we decomposed the total entropy into the logical entropy and average conditional entropy in Section 3.3, we can decompose the total mutual information into the correlation between the logical states and the average conditional mutual information between the physical states:

$$I_{in}(\Gamma_S;\Gamma_M) = I_H(S;M) + I_{in}(\Gamma_S;\Gamma_M|S \otimes M). \tag{3.80}$$

Taking the mutual information into account, the total change in entropy ΔS_{tot} after some arbitrary thermodynamic interaction between the system and the measurement apparatus is given by

$$\Delta S_{tot} = \underbrace{\Delta H^S + \Delta H^M - \Delta I_H}_{\text{logical entropy } \Delta H} + \underbrace{\Delta S_{in}^S + \Delta S_{in}^M - \Delta I_{in}}_{\text{internal entropy } \Delta S_{in}} - \underbrace{\beta Q}_{\text{heat}}, \tag{3.81}$$

where the superscript S and M indicates the system and entropy, respectively. Going back to the erasure process and Eq. (3.74), we see that if the internal entropy does not change during the erasure (i.e., the initial and final phase space distribution are equilibrium distributions), we obtain

$$\Delta H^S + \Delta H^M - \Delta I_H \geq \beta Q. \tag{3.82}$$

After the full cycle of measurement, expansion, and deletion of memory, the logical states of the system and the memory is the same as the initial ones. Therefore $\Delta H^S = \Delta H^M = 0$, and since the internal energy does not change we also have

$Q = -\Delta W$. Using this we obtain yet another version of Landauer's principle

$$W \geq \Delta I_H / \beta. \tag{3.83}$$

The work required to delete the information in a memory, is given by the mutual information between the system and the memory. In the case of a perfect measurement we have $\Delta I_H = \ln 2$, which means that the minimum work we have to pay to erase the memory is the same as the work we obtain from the Szilard engine.

3.4.1 Measurement errors

Measurement errors reduce the mutual information between the system and memory, and therefore the work required to delete the memory. However, as we argue in paper 2, it is not possible to saturate the bound in Eq. (3.83) when measurement errors are present. This is due to an irreversible entropy production not accounted for, which we will describe briefly in the following. If the system is a standard Szilard engine, and the memory is a single-particle-box as before, there are four distinct logical states $(00, 01, 10, 11)$. In Fig. 3.4 we show a schematic of the full phase space of the (system + memory). Here we reduce the dimension of the phase space to the only relevant degree of freedom (the x-coordinate). Therefore the horizontal axis represents the x-coordinate of the particle in the system, while the vertical axis represents the x-coordinate of the particle in the memory. The total phase space is divided into four quadrants, each of which represents one of four logical states, associated with which side of the box the particle is in the system and memory. The initial state of the system + memory is shown in Fig. 3.4(a), where the memory is in a standard state 0, while the system is either in the state 0 or 1 with probability 1/2. If an error-free measurement is performed on the system and copied into the memory, the full phase space evolve into what is shown in Fig. 3.4(b). The internal entropy and the logical state of the system and memory is identical; both the memory and the system is either in state 0 or 1 with the same phase space distribution.

Consider now the schematic in Fig. 3.5, showing the phase space evolution of this model when measurement errors are present. The initial state shown in Fig. 3.5(a), is the same standard state as in the error-free measurement. If the system is now put into contact with the measurement apparatus and copied into the

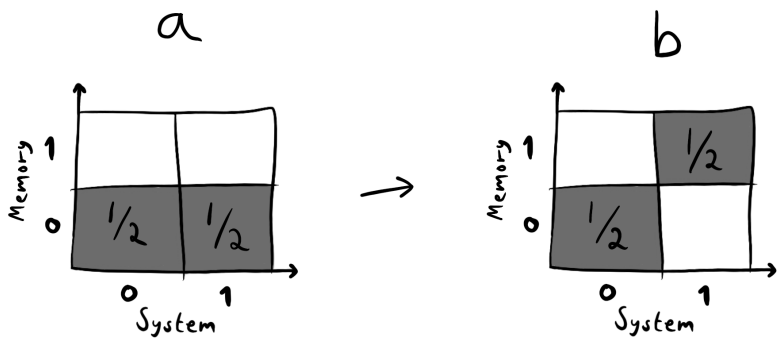

Figure 3.4: Illustration of the combined phase space of a Szilard engine and a single-particle-box memory. (a) shows the initial state where the memory is in the standard state 0 and the system is in either the 0 or the 1 state. The transition from (a) to (b) is an example of an error-free measurement, where the both the system and memory is either in the logical states 00 or 11.

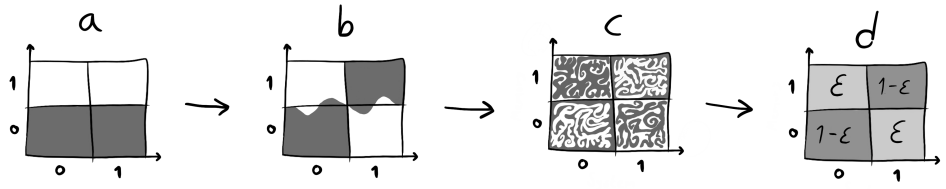

Figure 3.5: Illustration of the total phase space of a Szilard engine and a single-particle-box memory. (a) shows the initial state where the memory is in the standard state 0, while the system is in either 0 or 1 with probability 1/2. A measurement error occurs in (b1)/(b2), and once the barrier is inserted so the phase space can not flow between the quadrants, the phase space of the incorrectly mapped states evolve chaotically according to Hamiltonian dynamics as shown in (c). Coarse graining of the phase space after the time evolution results in the final phase space distribution shown in (d).

memory, some of the system states are incorrectly mapped to the memory. This incorrect mappings come from the cases where the actual position of the particle in the system does not agree with what was recorded in the memory, and is shown in Fig. 3.5(b), i.e., the phase space points in 01 are wrongly mapped and should fill in the empty space in 00. When the barrier is inserted the phase space points can no longer cross the boundaries between the four quadrants. However, the phase space continues to evolve according to deterministic Hamiltonian dynamics, resulting in a complicated structure of the phase space as shown in Fig. 3.5(c). Nevertheless, since the time evolution obeys Liouville's theorem, the entropy of Fig. 3.5(c) is still the same as in Fig. 3.5(b). To reach the final state with uniform phase space distributions, shown in Fig. 3.5(d), we have to coarse-grain the phase space. We therefore lose information about the complicated phase space structure. It is this coarse-graining that introduces an irreversible measurement entropy given by

$$S_\varepsilon = -\varepsilon \ln \varepsilon - (1-\varepsilon)\ln(1-\varepsilon), \tag{3.84}$$

where ε is the probability of measurement error.

3.4.2 Experimental detection

For a long time, Maxwell's demon has only been a thought-experiment, which was impossible to realize in the lab. However, in recent years, modern technology have enabled experimenters to create working versions of Maxwell's demon and close analogies, in a range of physical systems: atoms [35, 36, 37], colloidal particles [38, 39], molecules [40], electrons [41, 42, 43] and photons [44]. It is therefore natural to consider how theoretical predictions can be verified in these experimental set-ups. We have suggested a way to detect the measurement entropy, in an experimentally realized Szilard engine based on a single-electron-box [41], which we will present here.

The experimental set-up is a single-electron-box (SEB), consisting of two small metallic islands connected by a tunnel junction. The metallic islands contain an electron gas with a large number of electrons, but one is able to control the position of one extra electron. The potential of the islands can be controlled separately, and if the islands are in thermal equilibrium at some reference potential ($V_L = V_R = 0$), the probability to find the extra electron on either island is

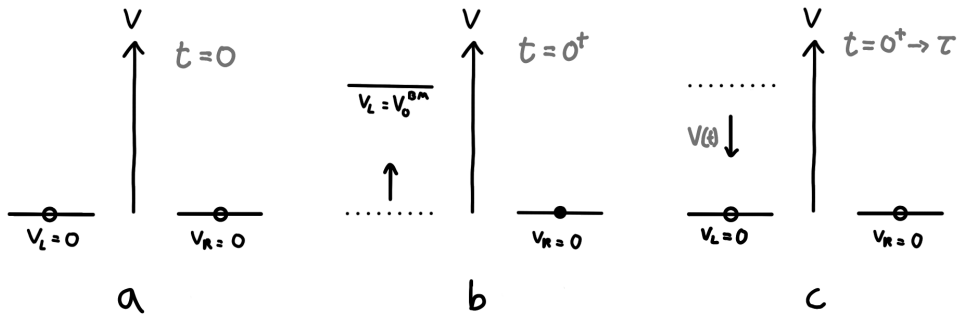

Figure 3.6: Two metallic islands share one excess electron. In (a) the islands are at equilibrium, with $V_L = V_R = 0$, with equal probability of finding the electron on either island. In (b), a measurement is performed, and the potential of the empty island is raised to some value V_0. As the potential of the island is gradually reduced back to zero, we can extract work whenever the electron occupies the island whose potential is being reduced, as shown in (c).

determined by the Boltzmann distribution

$$P_L = P_R = \frac{e^{-\beta V_{L/R}}}{e^{-\beta V_L} + e^{-\beta V_R}} = \frac{1}{2} \tag{3.85}$$

In Fig. 3.6 we show a schematic of the two islands, and how to perform the Szilard engine protocol for this system. Initially, at $t = 0$, the islands are at equilibrium, with $V_L(0) = V_R(0) = 0$, and there is a 50/50 chance to find the excess electron on either island, as shown in (a). A measurement is then performed, to determine which island the electron currently occupies. (b) If the electron is detected at the right island, we instantly raise the potential of the left island to some value $V_L(0^+) = V_0$. (c) We then slowly lower the potential of the left island back to zero at time, $V_L(0^+) = V_0 \to V_L(\tau) = 0$, where τ is the duration of the protocol. Due to thermal fluctuations, the excess electron jumps between the two islands, and whenever the electron occupies the right island while the potential is decreased, we extract an amount of work given by $W = \int_{t_0}^{t_1} eV(t)\, dV$, where e is the electron charge. If there are no measurement errors present in the experiment, and the measurement of which island we find the island is correct, the value of the initially raised potential, V_0, can be as large as we want. The larger the value, the more work we can extract as we lower it back to zero. However, if errors are present, we should limit the value of V_0. When a measurement error occurs, the island whose potential we raise is the one which has an additional electron, and we have

to expend a large amount of work, $W = eV_0$.

In our article paper 2, we found that in order to not produce any entropy in this step, the value we first raise the potential to, $V_L(t = 0^+) \equiv V_0$, has to be equal to the one we get from the equilibrium Boltzmann distribution for a given error rate ε,

$$P_R = \epsilon = \frac{e^{-V/k_BT}}{1 + e^{-V/k_BT}} \qquad \rightarrow \qquad V_0^{BM} = k_BT \log\left(\frac{1}{\epsilon} - 1\right), \tag{3.86}$$

This means that after you make a measurement and raise the potential of one of the islands, you are immediately in a new equilibrium state. If, after the measurement, the potential is raised to any other value, $V_0 \neq V_0^{MB}$, then there will be a relaxation from that initial distribution to the equilibrium Boltzmann distribution, with associated heat exchange.

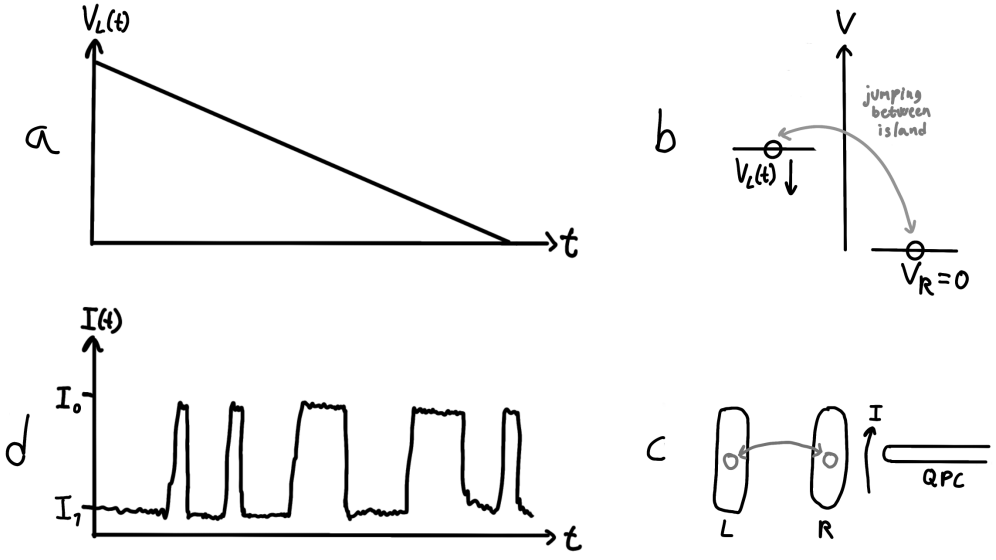

Figure 3.7: (a) A linear protocol for the potential of the left island as a function of time. (b) shows the electron jumping between the two islands due to thermal fluctuations. (c) illustration of how to detect which island the electron is on, by measuring the current through a quantum-point-contact (QPC). (d) the quantized QPC current as a function of time, which is can be mapped to (a) to find the energy of the electron as a function of time.

The way to experimentally detect which island the extra electron occupies, is to use a quantum-point-contact (QPC) probe. A schematic is shown in Fig. 3.7(c), where there are two metallic islands (L and R), and a QPC probe near the right island. There is a narrow channel between the right island and the QPC, where conductance is quantized [45]. Its conductance strongly depends on the local potential, so if there is an electron on the right island, the current through the channel I, is one unit of quantization lower than if it is no electron there. The current as a function of time, is shown in Fig. 3.7(d), where I_1 is the current when the electron is present on the right island, and I_0 is the current when it is not. In Fig. 3.7(a), we show the potential of the island as a function of time, for a linear protocol from $V_L(0^+) = V_0^{BM} \rightarrow V_L(\tau) = 0$. Since the QPC current gives us information about the position of the particle as a function of time, we can map this to the potential as a function of time, to find the time dependence of the energy of the extra electron. This plot is shown in Fig. 3.8, where the gray line $V(t)$ shows the potential of the unoccupied island we raise after a measurement, as a function of time. The red line shows the energy of the extra electron as a function of time. Whenever the electron jumps from the right island to the left island, an amount of heat Q_{in} is absorbed from the environment. Similarly, when the electron jumps from the left island to the right, an amount of heat Q_{out} is transferred to the environment. The amount of heat exchanged per jump, is given by $eV_L(t)$ at the time of the jump. Therefore, we can find the total heat exchange between the system and environment by summing up all these transitions

$$\bar{Q} = \sum_i Q_{in}^i - \sum_j Q_{out}^j, \tag{3.87}$$

The measured position of the particle is stored in a memory. When we reset the memory to complete a full cycle, we have to delete information by expending an amount of work W_D which has an associated heat released into the environment, Q_D, according to Landauer's principle. The heat dissipated via deletion can be found indirectly by the probability of finding the electron on either island at the time of measurement, and it will $Q_D = k_B T \ln 2$ if we set $V_L(\tau) = 0$ at the end of the protocol and allow time for equilibration between the islands before the next cycle starts.

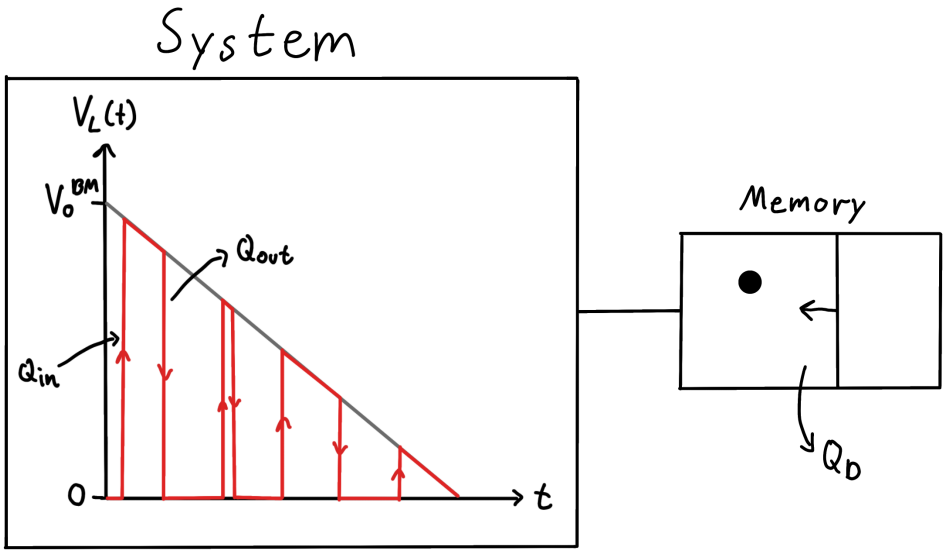

Figure 3.8: Schematic of the heat exchange between the environment, and the system and memory. The red line shows the energy of the electron as a function of time for a linear $V_L(t)$ protocol (gray line), which can be used to find the heat exchange between the environment and system. We can find the heat dissipation of erasing the memory by using Landauer's principle. If the experiment is repeated many times, and we find that it is equally likely that the electron occupies the particle on the left and right side, the cost of deleting the memory will be $Q_D = k_B T \ln 2$.

The net heat transferred between the environment and the system and memory is therefore

$$Q_{\text{net}} = Q_D - \bar{Q}. \tag{3.88}$$

This heat flow will result in a change of entropy of the universe, which is

$$S_{\text{tot}} = S_\varepsilon + \Delta S = \frac{Q_D - \bar{Q}}{T}. \tag{3.89}$$

Here S_ε is the entropy produced due to the measurement error, and ΔS is the entropy produced as the potential $V(t)$ is brought back to zero after the measurement. We predict that this error entropy can be found directly by measuring the error rate of the system and using the formula

$$S_\varepsilon = -\varepsilon \ln \varepsilon - (1-\varepsilon) \ln (1-\varepsilon). \tag{3.90}$$

However, to prove that the error entropy really exists, another way to find it (independent of our prediction) is by calculating the heat flow Q_{tot} as a function of power, $P = \bar{Q}/(k_B T \tau)$ (where τ is the duration of the protocol), and then calculate S_{tot}. Since ΔS will tend to zero as we approach the reversible regime (zero power), the intercept will be equal to S_ε as shown in Fig. 3.9.

$$\lim_{P \to 0} S_{\text{tot}} = \lim_{P \to 0} \frac{Q_D - \bar{Q}}{T} = S_\varepsilon. \tag{3.91}$$

A summary of the experimental procedure to detect the measurement entropy is given below.

1. Start with the initial condition being equal probability of occupying either of the two islands, with $V_L = V_R = 0$
2. Measure which island the particle is found, and raise its potential to $V_0^{BM} = \ln\left(\frac{1}{\epsilon} - 1\right)$, where ϵ is the error rate.
3. Adiabatically reduce the potential back to $V_L(\tau) = V_R(\tau) = 0$, while measuring the heat flow $\bar{Q} = Q_{in} - Q_{out}$.
4. When the potential is back to $V_L = V_R = 0$, allow the system to reach equilibrium by waiting for some time, and then repeat from 1.

As long as the lowering of the potential is performed adiabatically, the exact protocol to follow is not important, however a linear protocol is probably the simplest.

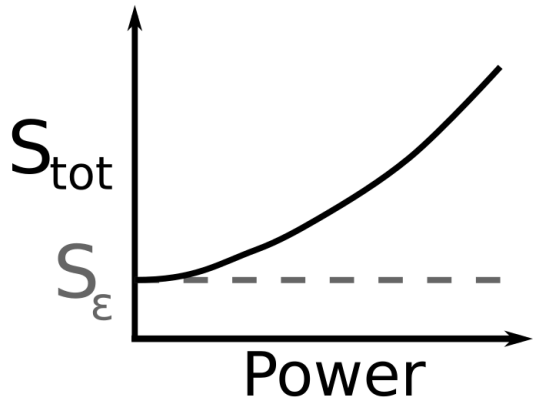

Figure 3.9: The total entropy S_{tot} as a function of power P. The intercept at zero power will be the measurement entropy S_ε.

Repeating this many times for a single duration of the protocol,τ, will give us statistics for $\langle \bar{Q} \rangle$. Also, the heat of deletion is given by $Q_D = \langle \sum_{i=L,R} P_L \ln P_L \rangle$, where P_L and P_R are the probabilities of finding the particle on the left and right island, respectively. These probabilities can be found by registering how many times the particle was found on the right/left island when measuring. If this is all repeated for decreasing values of τ, in the limit of $\tau \to \infty$, the total entropy production S_{tot} should approach the value $S_\epsilon = -\varepsilon \ln \varepsilon - (1-\varepsilon)\ln(1-\varepsilon)$, as shown in Fig. 3.9.

3.5 Asymmetric Szilard engine

In all the discussion about the Szilard engine so far, we always inserted the barrier in the center of the single-particle-box. If the barrier is inserted off center, such that there is an asymmetry between the compartment sizes, the situation is slightly changed.

Consider a box of width L as shown in Fig. 3.10. Initially the probability to find the particle anywhere is a uniform equilibrium distribution. If we now insert the barrier asymmetrically, such that width of the left and right compartment becomes

$$L_L = L(1-r), \quad \text{and} \quad L_R = Lr, \tag{3.92}$$

where, $r < 1/2$, is the ratio between the volumes of the right and left compart-

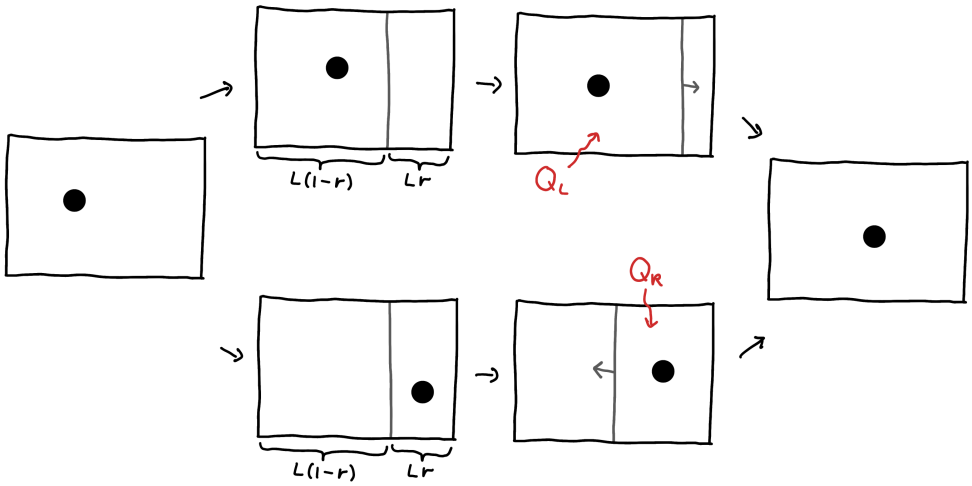

Figure 3.10: Schematic of the operation protocol for an asymmetric Szilard engine. Here r is the volume ratio of the two compartments after barrier insertion, L the total width of the box, and $Q_L(Q_R)$ the heat absorbed from the environment if the particle is detected on the left(right) side.

ment. The work we're able to extract from the engine during the isothermal barrier expansion depends on which side the particle is found when performing the measurement. The work depends on the initial and final volumes (V_i and V_f), in the following way: $W = k_B T \ln(V_f/V_i)$. Therefore we get

$$Q_L = -W_L = k_B T \ln r, \tag{3.93}$$

$$Q_R = -W_R = k_B T \ln(1-r). \tag{3.94}$$

Since $r < 1/2$, we see that $Q_R > Q_L$, and we're able to extract more work from the heat bath if the particle is detected in the smaller compartment (the one to the right in this case). However, the probability to find the particle in the smaller compartment is also lower. The average heat extracted over many cycles, is given by

$$\langle W_{exp} \rangle / k_B T = r \ln r + (1-r) \ln(1-r),$$

which is proportional to the change in Shannon entropy before and after measurement. This is exactly the same situation as for the symmetric Szilard engine; The average amount of work we can extract is given by the amount of information obtained by measurement. In this case however, the amount of information obtained

is smaller than in the symmetric case. Intuitively, this is because if you measure the particle in the left compartment, you are less surprised by this outcome, and this outcome also occurs more often. On the other hand, measuring the particle on the right side will surprise us more, but this event occurs less often.

The average work $\langle W_{exp} \rangle$ is maximized for $r = 1/2$, i.e. for a symmetric Szilard engine which nets us $W_{exp} = k_B T \ln 2$. Any value smaller than $r = 1/2$ will decrease the amount of work we're able to extract per cycle. Assume we have an asymmetric Szilard engine with $r < 1/2$. Since we're still storing this information in a binary memory, which we have to delete each cycle with associated work $W_{del} = k_B T \ln 2$, the total work per cycle is

$$\langle W_{tot} \rangle = \langle W_{exp} \rangle - \langle W_{del} \rangle < 0. \tag{3.95}$$

It seems that the average work per cycle is negative, which implies that the asymmetric Szilard engine performs much worse than the symmetric one. In fact, if we delete the memory every cycle, that is true.

However, instead of deleting the memory every cycle, we can keep a sequential record of the N previous measurement outcomes, and only delete the memory every N cycles. Since it is more probable to find the particle in the larger compartment (L), the data will contain uneven numbers of L and R measurement outcomes. The data can therefore be compressed before deletion. One example of such a compression strategy for the asymmetric Szilard engine is given in [46], where they report that after compressing a binary sequence of length N, the average work per bit needed to erase the compressed data is

$$\frac{\langle W_{del} \rangle}{N} = k_B T \left[r \ln r + (1 - r) \ln(1 - r) \right] + \mathcal{O}\left(1/\sqrt{N} \right), \tag{3.96}$$

which becomes exactly the same as $\langle W_{exp} \rangle$ when $N \to \infty$.

Chapter 4

Quantum information theory

In quantum systems, information processing in thermodynamic systems has subtle differences from the classical analogues. One of the main differences is that measurements in quantum mechanics are invasive; when a quantum state is measured, the wavefunction of the system collapses into a definite state. Hence, a measurement will change the state of the system that it is probing, which does not generally happen in classical systems. The purpose of this chapter is to give a complete picture of Maxwell's demon in both the quantum and classical picture, since papers 3 and 4 are based on quantum systems.

4.1 Basic introduction

The state of a quantum system is described by its wave function $|\Psi\rangle$. It can be written as a linear combination of any orthonormal basis $\{\,|n\rangle\,\}$

$$|\Psi\rangle = \sum_n c_n \,|n\rangle\,. \tag{4.1}$$

A measurement in quantum mechanics is based on the postulate that if you measure the system in a given basis, you are guaranteed that the system will be in one of these basis states $|n\rangle$. What sort of state the system was in before the measurement does not have a clear answer in quantum mechanics, other than that its mathematical description is given by the quantum superposition. According to the Born rule [47], the probability to find the system in a given state $|n\rangle$ is given by $|c_n|^2$. This type of measurement, where the outcome is always one of the basis states, is called a von Neumann measurement.

The wave function given in Eq. (4.1) is a pure state. There is no uncertainty regarding which state the system is in, even if it is described by a superposition of basis states; it is in the state $|\Psi\rangle$. In the cases where we might not know which state the system is in, we use the framework of density matrices. Similarly to the ensembles discussed previously, the density matrix is a probability density over all possible states $|\Psi\rangle$. While the coefficients c_n describes the inherent quantum mechanical uncertainty of a given state, the density matrix is a measure of our classical uncertainty about the system. The density matrix, describing our state of knowledge of a quantum mechanical system, where the probability of state $|\Psi_m\rangle$ is p_m, is defined as

$$\rho = \sum_m p_m \left|\Psi_m\right\rangle \left\langle\Psi_m\right| . \tag{4.2}$$

The expectation value of an observable A in terms of the density matrix is

$$\langle A\rangle = \sum_m p_m \left\langle\Psi_m\right| A \left|\Psi_m\right\rangle = \sum_m \mathrm{Tr}\left(A \left|\Psi_m\right\rangle \left\langle\Psi_m\right|\right) = \mathrm{Tr}\left(A\rho\right), \tag{4.3}$$

and the time evolution is given by

$$\rho(t) = U(t)\rho(0)U^\dagger(t), \tag{4.4}$$

where $U(t)$ is the time-evolution operator. The density matrix is the quantum mechanical equivalent of a classical probability density; If we express the set of possible states $\{\, |\Psi_m\rangle \,\}$ in the basis $\{\, |i\rangle \,\}$

$$|\Psi_m\rangle = \sum_n c_{im} \left|i\right\rangle , \tag{4.5}$$

the off-diagonal and diagonal matrix elements of ρ are given by

$$\rho_{ij} = \left\langle i\right| \rho \left|j\right\rangle = \sum_m p_m c_{im} c_{jm}^*, \tag{4.6}$$

and

$$\rho_{ii} = \left\langle i\right| \rho \left|i\right\rangle = \sum_m p_m |c_{im}|^2, \tag{4.7}$$

respectively. Here $|c_{im}|^2$ is the conditional probability that the system is in the basis state $|i\rangle$, given that the initial state was $|\Psi_m\rangle$. Therefore the diagonal elements gives the total probability of finding the system in the state $|i\rangle$, given the proba-

bility distribution $\{ p_m \}$ of initial states $\{ |\Psi_m\rangle \}$. A density matrix consisting of a single state $\rho = |\Psi\rangle \langle\Psi|$ is called a pure state. If the pure state is written in the basis $\{ |i\rangle \}$, we have

$$\rho = |\Psi\rangle \langle\Psi| = \sum_{ij} c_i c_j^* |i\rangle \langle j| , \tag{4.8}$$

which is a superposition of the basis states $\{ |i\rangle \}$. In general, a mixed density matrix is a probability distribution over different state vectors, as in Eq. (4.2). However, we can also consider the density matrix of the basis states

$$\rho = \sum_i p_i |i\rangle \langle i| , \tag{4.9}$$

which should be thought of as system being in one of the definite basis states $|i\rangle$, and the density matrix describing our uncertainty of which basis state it is in. This is in stark contrast to a superposition of basis states, where we cannot think of the system as being in one basis state or another. Since the density matrix is Hermitian ($\rho = \rho^\dagger$), it can always be diagonalized in some eigenbasis $\{ \lambda_i \}$

$$\rho = \sum_i \lambda_i |\lambda_i\rangle \langle\lambda_i| , \tag{4.10}$$

where λ_i are the eigenvalues. In this eigenbasis the system is never in a superposition of the eigenstates, only a statistical mixture.

In quantum statistical mechanics, the von Neumann entropy [48, 49] is the extension of the classical Shannon entropy. Given a quantum mechanical system with the density matrix ρ, its entropy is

$$S = -\text{Tr}(\rho \ln \rho). \tag{4.11}$$

We immediately see that for a diagonal density matrix, the von Neumann entropy is reduced to the classical Shannon entropy

$$S = -\text{Tr}\,(D \ln D) = -\sum_i \lambda_i \ln \lambda_i, \tag{4.12}$$

where λ_i are the eigenvalues. For a pure state density matrix the entropy is zero, while the maximum entropy of a finite dimensional Hilbert space is a maximally

mixed density matrix where all states are equally likely

$$\rho_{max} = \frac{1}{N} \sum_n |n\rangle \langle n| , \quad n \in [1, N]. \tag{4.13}$$

In that case the entropy becomes $S = \ln N$, where N is the number of states.

The density matrix for a system in thermal equilibrium with a heat bath at inverse temperature β, is a special temperature dependent mixed state, known as the Boltzmann state

$$\rho = \frac{1}{Z} \sum_k e^{-\beta E_k} |\Psi_k\rangle \langle \Psi_k| . \tag{4.14}$$

Here $|\Psi_k\rangle$ are the energy eigenvectors, E_k the corresponding energy eigenvalues, and Z is a normalization factor known as the partition function

$$Z = \sum_k e^{-\beta E_k} = \text{Tr}\left(e^{-\beta H}\right), \tag{4.15}$$

where H is the Hamiltonian of the system.

A von Neumann measurement in quantum theory can be described by a projection operators, $P_n = |n\rangle \langle n|$, associated with a given set of basis states $\{ |n\rangle \}$. If we perform a von Neumann measurement on a density matrix ρ, we will find that the system is in state $|n\rangle$ with probability $p_n = \langle n| \rho |n\rangle$. The density matrix, $\tilde{\rho}_n$ is the final density matrix after measurement, and to obtain it from the initial density matrix ρ we use the projection operators in the following way:

$$\tilde{\rho}_n = |n\rangle \langle n| = \frac{P_n \rho P_n}{\text{Tr}\left(P_n \rho P_n\right)}. \tag{4.16}$$

The denominator is there to rescale the density matrix so that the final state is properly normalized. Although the projection operators P_n are linear operators, the renormalization in the denominator makes the von Neumann measurement non-linear.

4.2 Work extraction for quantum Szilard engine

The main point of the classical Szilard engine is that the demon is able to utilize information obtained by measurement to extract useful work from a thermal

system. Classically, work implies applying forces that change the Hamiltonian of the system, e.g. a gas compressed by a piston. Conversely, heat is a change in the probability distribution of the energy spectrum, i.e. changing the temperature changes the Boltzmann distribution. For a quantum system with density matrix ρ and Hamiltonian H, the standard way [50, 51] of separating heat and work in the total energy change is

$$dE = d\text{Tr}\left(H\rho\right) = \text{Tr}\left(d(H\rho)\right) = \underbrace{\text{Tr}\left(dH\rho\right)}_{\text{work}} + \underbrace{\text{Tr}\left(Hd\rho\right)}_{\text{heat}}. \tag{4.17}$$

Therefore work is the change in energy due to the change in the energy eigenstates given by H, and heat is the change in the density matrix, i.e. the change in the distribution of the eigenstates. This separation into heat and energy is not completely satisfying, since work can be done on a quantum system which changes both the energy levels and their occupation probabilities, e.g. by rapidly changing the potential of a single-particle-box [52]. Nevertheless, this breakdown can be used to analyze the work extraction from a quantum Szilard engine with adiabatic barrier insertion.

There are several differences between the classical and quantum Szilard engine. Chief among them is the fact that no work is required to insert the barrier for the classical version, which is not the case for the quantum version. The insertion of the barrier can be modeled by increasing the height of a potential barrier in the center of the single-particle-box, which will change the energy levels, thus requiring input work. If the increase of the potential barrier is performed adiabatically, the probability distribution of the energy levels does not change, so $\Delta Q = 0$. The quantum work when isothermally changing an external parameter from X_i to X_f is then [50]

$$W = \beta^{-1} \sum_n \int_{X_i}^{X_f} \frac{\partial \ln Z}{\partial E_n} \frac{\partial E_n}{\partial X} dX \tag{4.18}$$

$$= \beta^{-1} \left[\ln Z(X_f) - \ln Z(X_i)\right] \tag{4.19}$$

where $Z = \sum_n e^{-\beta E_n}$ is the partition function. Thermalization between the system and the heat bath destroys all coherence between the energy states, so the density matrix becomes diagonal with diagonal matrix elements $p_n = \exp\left(-\beta E_n\right)$. The usual process of wall insertion, measurement, and wall expansion for the quantum Szilard engine can be described by the following:

After the insertion of the potential barrier wall in the center of a single-particle-box, but before the measurement is performed, the partition function is

$$Z_{ins} = Z_0(L/2) + Z_1(L/2), \tag{4.20}$$

where Z_0 and Z_1 are the partition function for the case where the particle is found on the left and right side of the barrier, respectively. They are given by

$$Z_0 = Z_1 = Z(L/2) = \sum_n e^{-\beta E_n}, \quad E_n(L/2) = \frac{\pi^2\hbar^2}{2m(L/2)^2}n^2, \tag{4.21}$$

where L is the total width of the box. The initial partition function is

$$Z_{init} = Z(L) = \sum_n e^{-\beta E_n}, \quad E_n(L) = \frac{\pi^2\hbar^2}{2m(L)^2}n^2, \tag{4.22}$$

and therefore the amount of work required for the insertion of the barrier is

$$W_{ins} = \beta^{-1}\left(\ln Z_{ins} - \ln Z_{init}\right) \tag{4.23}$$

$$= \beta^{-1}\left(\ln 2 - \ln \frac{Z(L)}{Z(L/2)}\right) \tag{4.24}$$

In the low temperature limit, the particle will be in the ground state E_1 with unit probability and the work becomes

$$W_{ins} = \beta^{-1}\left(\ln 2 - [E_1(L) - E_1(L/2)]\right) \tag{4.25}$$

A measurement is performed to find which side of the barrier the particle is found, and the wall is allowed to expand into the empty side. After the expansion of the wall, the particle occupies the full volume of the box again, and the partition function becomes

$$Z_{exp} = Z(L) \tag{4.26}$$

Regardless of which side the particle is found, the amount of work that can be

extracted during the isothermal process is given by

$$W_{exp} = \beta^{-1}\left(Z_{0/1} - Z_{exp}\right) = \beta^{-1}\ln\frac{Z(L/2)}{Z(L)} \tag{4.27}$$

We see that the total work is given by

$$W_{tot} = W_{ins} + W_{exp} = k_B T \ln 2. \tag{4.28}$$

Since this is a positive value, it means we can extract an amount $k_B T \ln 2$ of work, as in the classical case. In the limit of $\beta \to 0$, we recover the classical result, where $W_{ins} = 0$ and $W_{exp} = k_B T \ln 2$.

4.3 Quantum measurement

In the previous section we showed how to extract a work amount of $W_{tot} = k_B T \ln 2$, from a Szilard engine. In this section we discuss how a quantum demon gains information on the state of a quantum Szilard engine, following [53]. As in the previous section, we consider the Szilard engine to be a single-particle box, which after the barrier insertion is in the state $|L_n\rangle$ or $|R_n\rangle$, where n is a positive integer. Since we have a binary measurement outcome, we can model the quantum demon as a two-state quantum system. Assume that the demon is in an initially prepared state $|D_0\rangle$. Since we only measure which side the particle is on, and not its exact eigenstate, the measurement process is accomplished by the transition

$$|L_n\rangle|D_0\rangle \;\to\; |L_n\rangle|D_L\rangle \tag{4.29}$$

$$|R_n\rangle|D_0\rangle \;\to\; |R_n\rangle|D_R\rangle\,. \tag{4.30}$$

Since we want the measurement to be unambiguous, $|D_R\rangle$ and $|D_L\rangle$ has to be orthogonal. An example of an interaction Hamiltonian between the system and demon that accomplishes this is given in [54]. That is, an interaction

$$H_{int} = i\delta\left(|L_n\rangle\langle L_n| - |R_n\rangle\langle R_n|\right)\left(|D_L\rangle\langle D_R| - |D_R\rangle\langle D_L|\right), \tag{4.31}$$

acting for a duration $\Delta t = \pi\hbar/4\delta$, when the initial state of the demon is

$$|D_0\rangle = \left(|D_L\rangle + |D_R\rangle\right)/\sqrt{2}, \tag{4.32}$$

and the initial state of the engine (after barrier insertion) is

$$\rho_E = Z^{-1} \sum_k^{\infty} e^{-\beta E_n(L/2)} \left(|L_n\rangle \langle L_n| + |R_n\rangle \langle R_n| \right), \tag{4.33}$$

results in a full density matrix for the joint system of demon and engine

$$\rho = e^{-iH_{int}\Delta t/\hbar} \rho_E |D_0\rangle \langle D_0| \simeq \left(\rho_L |D_L\rangle \langle D_L| + \rho_R |D_R\rangle \langle D_R| \right)/2, \tag{4.34}$$

where

$$\rho_L = Z^{-1} \sum_{n=1}^{\infty} e^{-\beta E_n(L/2)} |L_n\rangle \langle L_n|, \tag{4.35}$$

$$\rho_R = Z^{-1} \sum_{n=1}^{\infty} e^{-\beta E_n(L/2)} |R_n\rangle \langle R_n|. \tag{4.36}$$

If we now take the partial trace of the joint density matrix ρ, to find the density matrix of the engine after the interaction we obtain

$$\bar{\rho}_E = \mathrm{Tr}_D(\rho) = \langle D_L| \rho |D_L\rangle + \langle D_R| \rho |D_R\rangle = \rho_E, \tag{4.37}$$

and therefore the entropy, $S(\rho) = -k_B \mathrm{Tr}(\rho \ln \rho)$, of the gas is unchanged for an external observer. On the other hand, the state of the demon will have changed from an initial pure state $|D_0\rangle \langle D_0|$ to a mixture

$$\bar{\rho}_D = \mathrm{Tr}_E(\rho) = \sum_n^{\infty} \left(\langle L_n| \rho |L_n\rangle + \langle R_n| \rho |R_n\rangle \right) = \left(|D_L\rangle \langle D_L| + |D_R\rangle \langle D_R| \right)/2, \tag{4.38}$$

and therefore the entropy of the demon changes by an amount

$$\Delta S = S(\bar{\rho}_D) - S(|D_0\rangle \langle D_0|) = k_B \ln 2. \tag{4.39}$$

So the sum of the entropies of the demon and engine has increased, yet the entropy of the joint system could not have increased: the measurement process was a dynamical quantum time-evolution on an isolated engine-demon system, and unitary operators conserve information. The answer is that the loss of information regarding the demon, is compensated by an equal increase in the mutual

information between the demon and the engine

$$\Delta I = I(\rho) - I(\rho_E \otimes \rho_D) = k_B \ln 2. \tag{4.40}$$

From the conservation of information, it follows that an increase in entropy must be accompanied by an equal increase in the mutual information

$$\Delta I = \Delta S. \tag{4.41}$$

This mutual information can be used to extract $k_B \ln 2$ amount of work from the engine, as discussed in the previous section. However, after this work extraction, the demon is still in a mixed state, and to perform the next cycle of measurement and work extraction the state of the demon must be reset. By leaving the demon in a mixed state when initiating the next cycle, the coupling with the engine via the interaction H_{int} would not result in an increase of the mutual information. In order to reset the joint system to its initial state, we have to delete the information that is already stored in the demon. How to do this will be discussed in the next section.

4.4 Quantum Landauer's principle

We are now equipped with the tools needed to formulate Landauer's erasure principle for quantum states, and we will follow the derivation in [49]. We can model the standard binary memory as a two state system, which is either in the state $|0\rangle$ or $|1\rangle$. The memory is coupled to a finite sized bath, which is in one of its N accessible microstates. The N accessible microstates constitutes a subspace of the full Hilbert space of the bath. Since the state of the bath is inaccessible to us, as the observer, we describe it by a maximal entropy state, where all of its states are equally likely

$$\rho_{bath} = \frac{1}{N} \sum_n |n\rangle \langle n| . \tag{4.42}$$

Since the state of the bath is a maximum entropy mixed state, its entropy is

$$S(\rho_{bath}) - k_B \mathrm{Tr} \left[\rho_{bath} \ln \rho_{bath}\right] = k_B \ln N \tag{4.43}$$

The initial state of the memory is $\rho_{mem} = |i\rangle\langle i|$, where i is either 0 or 1. The initial state of the full system (memory + bath) is therefore

$$\rho_0^{init} = |0\rangle\langle 0| \otimes |n\rangle\langle n|, \quad \text{or} \quad \rho_1^{init} = |1\rangle\langle 1| \otimes |n\rangle\langle n|, \tag{4.44}$$

where $|n\rangle$ is the unknown state of the bath. Let's consider the $|0\rangle$ state as the standard state of the memory, which it is reset to after deletion. The final state of the total density matrix after deletion is then

$$\rho_i^{fin} = U\rho_i^{init}U^\dagger = |0\rangle\langle 0| \otimes \rho_{bath}^i, \tag{4.45}$$

where i can be either 0 or 1. Since the time evolution U operator is unitary, all joint system-bath initial states, $|i\rangle|n\rangle$, that were once orthogonal remains orthogonal in the final states they are mapped to. This is equivalent to logical reversibility: by applying the inverse time evolution on the final state, we end up in the same initial state. The joint state of the memory and bath, $|0\rangle|n\rangle$ is therefore mapped to orthogonal states, and the final density matrix is on the form

$$\rho_i^{fin} = |0\rangle\langle 0| \otimes |n,i\rangle\langle n,i|, \tag{4.46}$$

where $U|i\rangle|n\rangle = |0\rangle|n,i\rangle$. Now since the two possible initial states of the joint system, ρ_0^{init} and ρ_1^{init} are orthogonal, the two states ρ_0^{fin} and ρ_0^{fin} will also be orthogonal. Since $|0\rangle$ and $|1\rangle$ are orthogonal, we see from Eq. (4.46) that this implies that the states $|n,0\rangle$ are all orthogonal to the states $|n,1\rangle$

$$\langle n,0|m,1\rangle = 0, \quad \forall \quad n,m. \tag{4.47}$$

This is a statement of reversibility: the set of states $|n,0\rangle$ and $|n,1\rangle$ can not have any states in common, since if that was the case, we would not be able to discover what the initial state was by measuring the final state.

Now, we have to impose the key condition of erasure; the observer should not be able to determine which state the system was in initially. Therefore the final microstates of the bath, $|n,0\rangle$ or $|n,1\rangle$, must correspond to the same macroscopic state. Since the initial macrostate of the bath contains N microstates, the final state of the bath must therefore be a macrostate containing $2N$ microstates. In other words, the accessible microstates (which is still a subspace of the full Hilbert space) of the bath, have increased from N to $2N$. The change in entropy

is therefore

$$\Delta S = -k_B T \ln 2N + k_B T \ln N = k_B T \ln 2, \tag{4.48}$$

which is the same as the classical case. The Landauer principle has recently been confirmed in a fully quantum system [55].

Chapter 5

Deep reinforcement learning

Deep Reinforcement Learning is an exciting field, based on machine learning. It is a highly adaptable technique, that can be used to solve a large variety of tasks. In paper 4, we use this technique to solve an optimization problem related to a quantum Szilard engine. Therefore, we give a general introduction to the field in this chapter, as well as explain the specific algorithm we used in detail. It is not required to know anything about deep learning to understand this chapter, and we will not go into technical details of neural networks. While reading this chapter, one can essentially think of the networks as function approximators. For more information on the basics of neural networks and why they are such excellent function approximators, see [56, 57].

5.1 Short introduction to reinforcement learning

There are three main subfields in modern machine learning: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning can be summarized as training a neural network to learn a function that maps input data to some output. The function is inferred by comparing the current output of the network to the desired output and updating the network parameters. Unsupervised learning is used to find commonalities in the input data, and is used to group, classify or categorize unlabeled data.

Reinforcement learning differs from supervised and unsupervised learning and is based on letting an *agent* learn how to behave in a desired way by taking actions in an *environment* and observing the effect of the action on the environment.

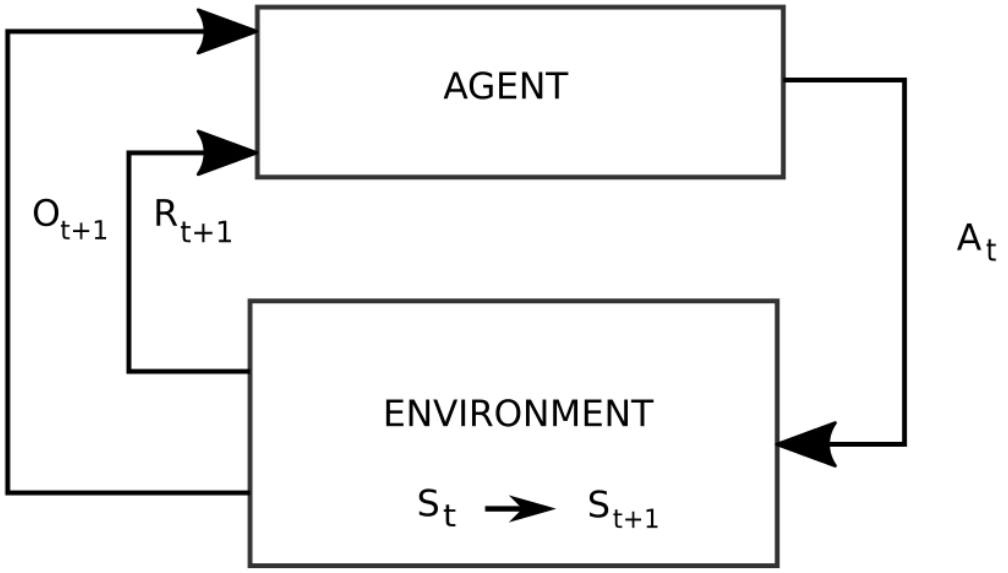

Figure 5.1: Schematic showing the basic formulation of the reinforcement learning. An agent performs an action A_t which induces a state change of the environment from S_t to S_{t+1}. The agent then receives an observation of the new state of the environment, O_{t+1}, and a reward, R_{t+1} that tells it how good the previous action was.

In order to define the "optimal" behavior of the agent, we give it feedback in the form of a *reward* based on the effect of its previous action. If the action change the environment into a more desirable state we give it a positive reward, while if it had negative consequences we give it a negative reward. A schematic of the basic reinforcement learning protocol is shown in Fig. 5.1. At time t the environment is in a given state S_t. The agent performs an action A_t which induces a state change of the environment from S_t to S_{t+1}. The agent then receives an observation, O_{t+1}, of the new state of the environment. This observation may be an observation of the full state, i.e. $O_{t+1} \equiv S_{t+1}$, or it can be a partial observation such that it is a subset of the full state, i.e. $O_{t+1} \subset S_{t+1}$.

As an example task that's suitable for reinforcement learning, consider the archetypal the cart pole balancing problem. An inverted pendulum is attached to a cart as shown in Fig. 5.2. The task is to balance the inverted pendulum in its upright position by moving the cart right or left. In this task, the state of

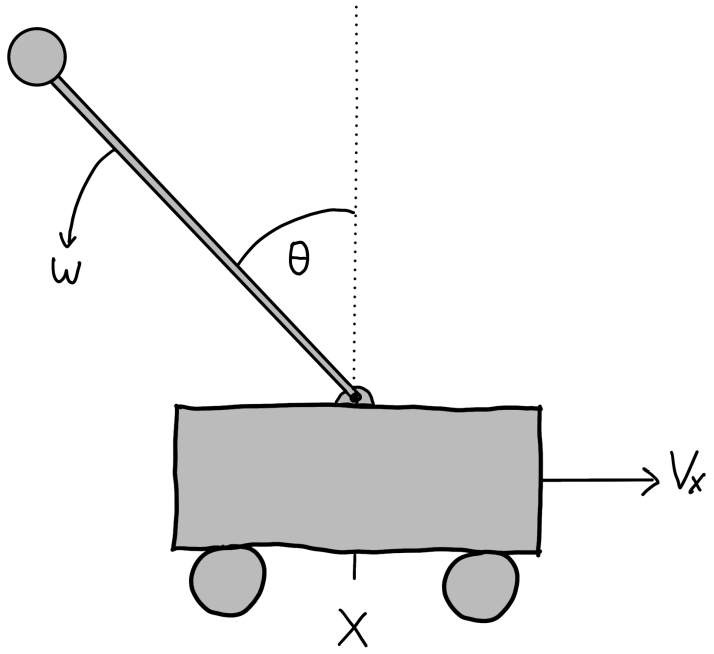

Figure 5.2: Schematic of the cart pole system. The goal is to minimize the angle θ, and the state of the system is fully described by the position x, horizontal velocity v_x, the angle of the pole θ, and its angular velocity ω.

the environment would be described by the position of the cart x, its velocity v_x, the angle of the pendulum with respect to the central axis θ, and the angular momentum of the pendulum ω. The state space would then be $S \in \{x, v_x, \theta, \omega\}$. Usually the more of a state the agent is allowed to observe, the easier it is to learn the desired behavior, so let's say the agent's observation is complete $O \equiv S$. The reward could be $r = +1$ for every timestep the pole has not fallen below a certain angle, and $r = -10$ if the pole falls down. The reward space is then $R \in \{+1 \vee -10\}$. In the simplest case, the actions available to the agent could be to apply a certain amount of force, F_x, in the x-direction to either the left or right side of the cart. The action space is therefore $A \in \{-F_x \vee +F_x\}$. In order to understand how we could utilize this information to teach the agent to balance the inverted pendulum we first need to understand the mathematical foundation of modern reinforcement learning; Markov Decision Processes.

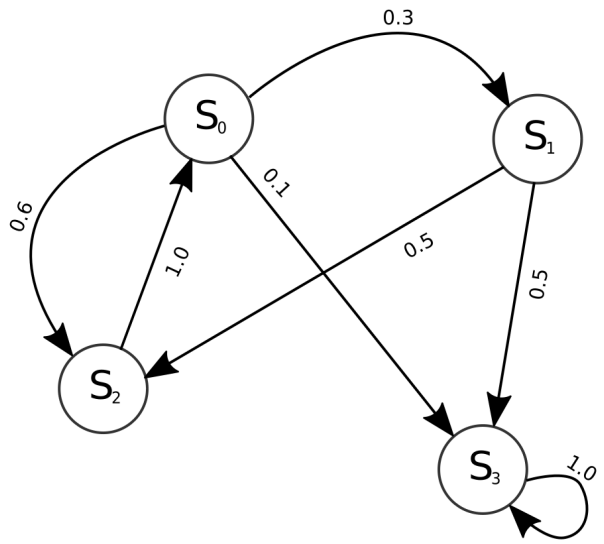

Figure 5.3: An example of a simple Markov chain model. This one have 4 different states, and the transition probability between each state is indicated by the numbers next to the arrows.

5.2 Markov decision process

5.2.1 Basic introduction

A Markov Chain is a stochastic model for the transitions between different states, where the transition probability only depends on the current state. The independence of the future state on the full history of the states visited is called the Markov property. Mathematically it can be defined as

$$P(s_{t+1}|s_t) \equiv P(s_{t+1}|s_t, s_{t-1}, \ldots, s_0), \tag{5.1}$$

i.e., the conditional probability for transitioning to the state s_{t+1} from the state s_t does not depend on the preceding states $s_{t-1}, s_{t-2}, \ldots, s_0$. In Fig. 5.3 a simple Markov chain model is shown. It consists of four different states, S_0, S_1, S_2, and S_3. The numbers next to the arrows indicate the transition probability from each state, and the sum of the transition probabilities from each state is equal to 1. In this example the state S_3 is a terminal state, since the only transition available is to itself. Markov decision processes was introduced by Richard Bellman in 1957 [58]. An MDP is similar to a Markov chain, but now transitions between states

are mediated by the choice of a set of actions available at each state. In addition, some state transitions result in a reward, which can be either negative or positive. An example of a Markov decision process is shown in Fig. 5.4. Here the actions are chosen by the agent, resulting in a stochastic transition to other states. In the state S_0 there are three available actions to choose from, A_1, A_2 and A_3. After choosing the action A_3 there is a 10% chance to move to state S_2, receiving a reward of $R = +100$, and a 90% chance to move to state S_3, receiving $R = +10$. Assuming we ended up in state S_2 there is only one available action, A_1, which have $50\%/50\%$ probability to take us either back to S_0 or to the terminal state S_3. Once we reach state S_3 there is again only one available action, which takes us back to S_3 with a reward (punishment) of $R = -10$. In this example the transitions after choosing an action are stochastic, but a MDP can also be deterministic, i.e. a given action in a given state always result in the same transition.

The goal of the agent is to find a policy (what actions to take at each state), which maximizes the total reward received. In the example given one might be able to find the optimal policy by inspection, but for larger more complicated MDPs this approach quickly becomes impossible. Richard Bellman found a way to estimate the optimal policy of a MPD, but to understand it we first need to define the value and quality functions.

5.2.2 Value function and Quality function

Consider the 2D grid word example shown in Fig. 5.5. Here the state of the system is given by the coordinates of the grid, e.g. $s = (1, 1)$ is the state in the upper left corner. There are two special states in this example; if $s = (1, 4)$ we get a reward of $R = +10$, while for $s = (3, 4)$ we get a reward of $R = -10$. Both of these states are also terminal states, so if we reach that state the game is over. The state $s = (2, 2)$ is also in a sense special, because it is unattainable. For each state there are four possible actions available to the actor; it can move in either of the cardinal directions. If the agent moves into any wall, it ends up in the same state it started in. A *policy* (denoted $\pi(a|s)$) in RL is an instruction to how the agent should act for any given state. On the left side of Fig. 5.5 the possible actions available for each state is indicated by the arrows. This is a random policy, and if the goal is to get to the state that gives you a reward of $R = +10$, it is not a very good policy. In this example, the optimal policy is easy to find by inspection, and it is shown on the left side of Fig. 5.5.

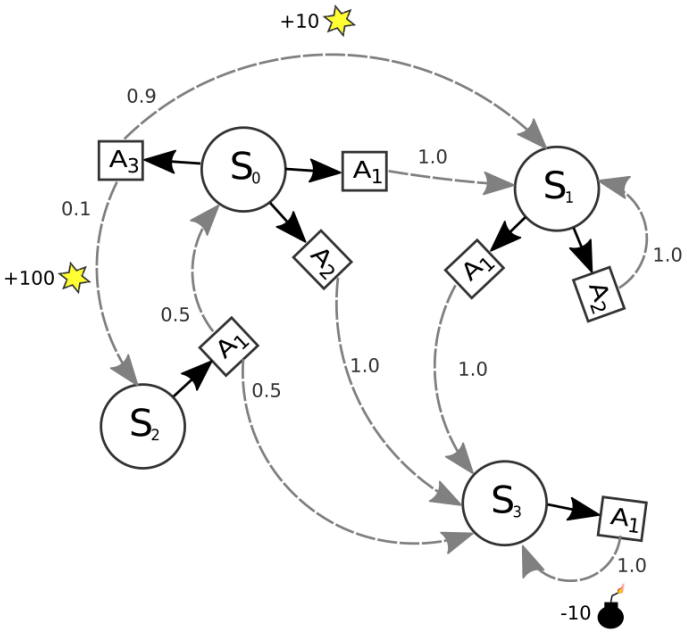

Figure 5.4: An example of a Markov decision process. As in the Markov chain model there are four possible states, but now each state has between one and three possible actions available. Some actions also result in negative or positive rewards.

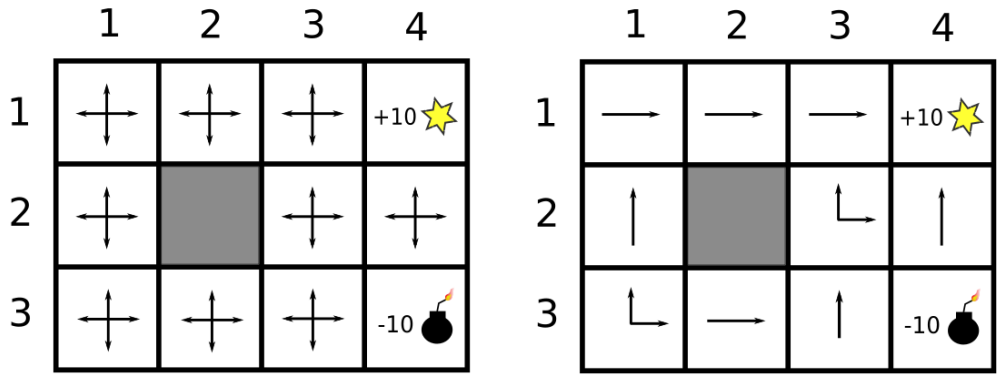

Figure 5.5: A 2D gridworld with $4 \times 3 - 1$ possible states and two terminal states, $(3, 4)$ and $(1, 4)$ which give rewards $+10$ and -10 respectively. The arrows indicate the actions available at each state. On the left side we have a random policy, and on the right side we show the optimal policy.

The value function $V_\pi(s)$ for a given policy π is just the expected cumulative reward gained by following the policy from the state s onwards. The value function for the optimal policy of the grid word example is shown on the left side of Fig. 5.6. Since the maximum reward we can get is $R = +10$, and the optimal policy we found always takes us to this state, the value function is $V_\pi(s) = 10$ for all states except the state $s = (3, 4)$ which have the value $V_\pi(3, 4) = -10$. An important concept we have to introduce now is the discount factor γ. Because of inflation and the possibility to gain interest on bank deposits, receiving 100 \$ now is better than receiving 100 \$ later, and the discount factor is meant to account for situations where this concept is applicable. For the cart-pole example from the previous section, the actions the agent have recently performed are more important to stabilize the inverted pendulum than actions it performed 100 timesteps earlier. Applying a discount factor of $\gamma = 0.9$ to the value function of the gridworld example gives us the value function as shown on the right side of Fig. 5.6.

The Q-function (quality function) is closely related to the value function, the only difference is that it gives the expected cumulative reward given a state-action pair. $Q_\pi(s, a)$ gives you the expected cumulative reward given that you are in state s, take the action a and follow the policy thereafter. The Q-function for the gridworld example with a discount factor of $\gamma = 0.9$ is shown in Fig. 5.7. The grids are now divided in four, one triangle for each possible action, and the brightness of the color illustrates the Q-function value for each state-action pair.

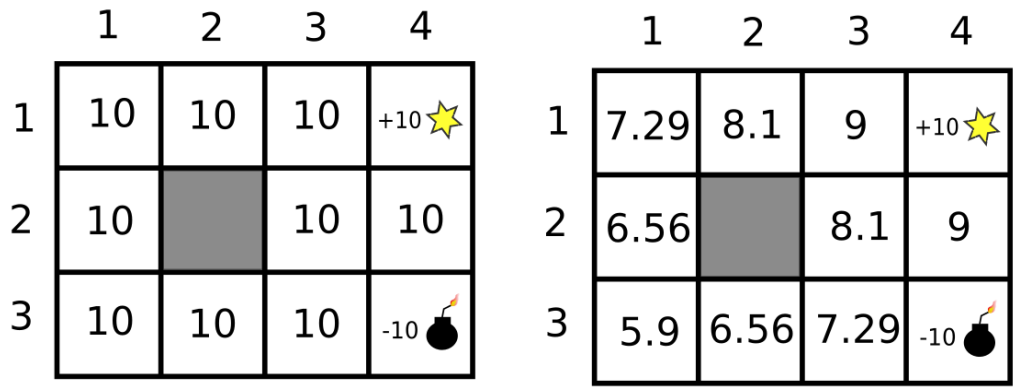

Figure 5.6: In this figure we show the same gridworld as above, only now we have filled in the value of each state, given that we follow the optimal policy. The right one is shown without a discount factor, and on the left we have applied a discount factor of $\gamma = 0.9$.

For a given state, actions that take us away from the goal are worse than actions that take us towards it, and the total quality of a state increases progressively as we get closer to the goal. The actions that take us to the negative terminal state are obviously the worst possible, so they are indicated by a red color. The advantage of describing the system by the Q-function instead of the value function is that the former encodes both the value of being in a certain state, and the policy to follow. So by finding the optimal Q-value of all the state-action pairs, denoted $Q^*_\pi(s, a)$, we also find the optimal policy $\pi^*(a|s)$. If we always choose the brightest shade of green in each state of Fig. 5.7 we see that they correspond to the arrows of the optimal policy shown in Fig. 5.5.

5.3 Deep Q-Learning

5.3.1 Basic formalism

So how do we find the optimal Q-value? The value function can be defined as

$$V_\pi(s_t) = E\left[R(s_t, a_t, s_{t+1}) + \gamma R(s_{t+1}, a_{t+1}, s_{t+2}) + \cdots + \gamma^n R(s_{\tau-1}, a_{\tau-1}, s_\tau)\right], \tag{5.2}$$

where $R(s_t, a_t, s_{t+1})$ is the reward received by going from state s_t to s_{t+1} via the action a_t, and s_τ is a terminal state. The discount factor is applied for all the future rewards and $E[\,]$ indicate the expectation value, in case there is stochasticity in the

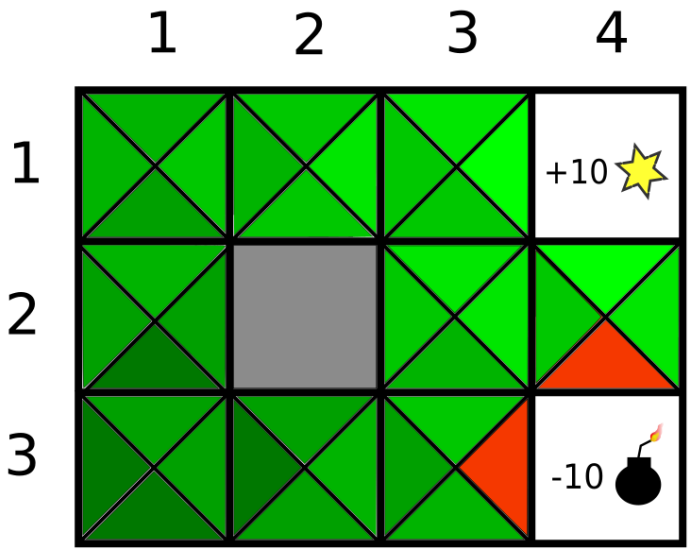

Figure 5.7: Here we illustrate the quality function of the grid word example with a discount factor of $\gamma = 0.9$. The brightness of the green color indicates how good the state-action pair is (brighter means higher expected reward).

transitions. This equation can be rewritten as

$$V_\pi(s_t) = E\left[R(s_t, a_t, s_{t+1}) + \gamma V_\pi(s_{t+1})\right], \tag{5.3}$$

since

$$V_\pi(s_{t+1}) = E\left[R(s_{t+1}, a_{t+1}, s_{t+2}) + \cdots + \gamma^{n-1} R(s_{\tau-1}, a_{\tau-1}, s_\tau)\right]. \tag{5.4}$$

We define the transition probability $T(s_t, a_t, s_{t+1})$ as the probability to go from state s_t to s_{t+1} given that we choose the action a_t, and write the expectation explicitly as

$$V_\pi(s_t) = \sum_{s_{t+1}} T(s_t, a_t, s_{t+1}) \left[R(s_t, a_t, s_{t+1}) + \gamma V_\pi(s_{t+1})\right]. \tag{5.5}$$

If we have found the optimal Value function $\pi^*(a_t|s_t)\ \forall\ t$ the following relation is obviously true

$$V^*_\pi(s_t) = \max_{a_t} \sum_{s_{t+1}} T(s_t, a_t, s_{t+1}) \left[R(s_t, a_t, s_{t+1}) + \gamma V^*_\pi(s_{t+1})\right]. \tag{5.6}$$

That is, the optimal Value function (maximum expected future reward) is obtained by taking the action a_t that maximizes the expected immediate reward obtained plus the expected reward from all possible future states that this action leads to. This equation is known as the Bellman Optimality Equation [56]. It has a very similar form for the Q-value function;

$$Q^*_\pi(s_t, a_t) = \max_{a_t} \sum_{s_{t+1}} T(s_t, a_t, s_{t+1}) \left[R(s_t, a_t, s_{t+1}) + \gamma \max_{a_{t+1}} Q^*_\pi(s_{t+1}, a_{t+1}) \right]. \tag{5.7}$$

This equation can be used to iteratively update the estimated value for the optimal Q-value for every possible state-action pair:

$$Q^{k+1}_\pi(s_t, a_t) \leftarrow \sum_{s_{t+1}} T(s_t, a_t, s_{t+1}) \left[R(s_t, a_t, s_{t+1}) + \gamma \max_{a_{t+1}} Q^k_\pi(s_{t+1}, a_{t+1}) \right]. \tag{5.8}$$

Here $Q^k_\pi(s_t, a_t)$ is the estimate of the Q-value for the kth iteration of the algorithm. These estimates are guaranteed to converge to the optimal Q-value, given enough iterations [56]. Once the optimal Q-value is found, the optimal policy is given by

$$\pi^*(a_t|s_t) = \operatorname*{argmax}_{a_t} Q^*(s_t, a_t) \tag{5.9}$$

This is an example of dynamic programming; we break down the complex problem of directly finding the optimal policy to the subproblems of finding the optimal Q-value for each state-action pair, which we then use to extract the optimal policy.

Finding the optimal policy using Eq. 5.8 can certainly be effective for small systems. However, the algorithm scales very poorly for larger MPDs with many states and actions. Using it we have to calculate values for all the state-action pairs. This is certainly possible for the grid word example introduced earlier, where we have approximately $\sim (3 \times 4) \times 4 = 48$ state-action pairs, but for the inverted pendulum the state-space is continuous, so we have in principle an infinite number of state-action pairs. The solution was introduced by DeepMind in 2015 [59], and involves approximating the Q-function using a neural network.

$$Q(s, a) \simeq Q(s, a, \theta). \tag{5.10}$$

Here θ are the parameters of the neural network (its weights and biases). Neural networks are excellent function approximators, and with this innovation, Deep-

Mind was able to greatly outperform humans in several Atari games [59]. Training a neural network to approximate the Q-values is called *Deep Q-Learning* (DQL).

A neural network trains by minimizing a loss function (also called a cost function). Let's say you want to predict housing prices in Oslo by looking at features such as the latitude and longitude, number of bedrooms, age of the house, and so on. You have obtained a certain amount of data where you have the prices of the houses as well as those features. A common loss function is the mean squared error (MSE)

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2, \tag{5.11}$$

where n is the number of samples, y_i is the true housing price (the target value) and $\hat{y}_i$ is the price predicted by the neural network. The MSE of the output from a neural network can be minimized by calculating its gradient and adjusting the weights and biases of the network accordingly. When the gradient of the MSE is zero we have reached a minimum (hopefully a global minimum) of the loss function. The details on the special back-propagation method used for training neural networks can be found in several articles, blog posts, and books, including [57].

When training the network in DQL, we do not actually know the real optimal Q-value. However we know that according to the Bellman equation the optimal Q-value satisfies

$$Q^*(s_t, a_t, \theta) = E\left[r_t + \gamma \max_{a_{t+1}} Q^*_\pi(s_{t+1}, a_{t+1}, \theta)\right], \tag{5.12}$$

where we have introduced the short-hand notation $r_t = R(a_t, s_t, s_{t+1})$. We can use this estimate as our target when training the network. The MSE loss function for DQL then becomes

$$L(\theta) = E\left[\left(r_t + \gamma \max_{a_{t+1}} Q^*_\pi(s_{t+1}, a_{t+1}, \theta) - Q(s_t, a_t, \theta)\right)^2\right] \tag{5.13}$$

5.3.2 Improvements

For DQL to work properly and converge to a good estimate of the optimal Q-values, there are several improvements we can make. Some of these were introduced in [59], and will be covered in the following.

Exploration vs Exploitation

When the network is initialized its predictions for the Q-values are of course totally wrong. So if we always chose the actions that maximizes the current predicted Q-values, $\underset{a_t}{\text{argmax}}\, Q^*(s_t, a_t)$, the agent would not learn anything. We need to let the agent explore the state-action space by randomly performing actions. A typical exploration policy is the ϵ-greedy policy. In the beginning of the training the agent chooses random actions with probability ϵ, or the ones with the highest Q-value (greedily) with probability $1 - \epsilon$. As time goes and the agent explores more and more of the environment, ϵ is decreased so that it focuses more on the areas of the state-action space with higher Q-values. Typically, we start by taking completely random actions, $\epsilon = 1$, and let ϵ converge to some finite number $\epsilon \sim 0.05$, so that there is always some exploration going on.

Experience Replay

As seen in Eq. 5.13 a single update of the network weights requires the following input: the current state s_t, the action chosen a_t, the immediate reward r_t, and the next state s_{t+1}. We call this tuple, $e_t = (s_t, a_t, r_t, s_{t+1})$, that the network trains on an experience. Instead of training on consecutive experiences we store them all in a memory $M_t = \{e_0, e_1, \ldots, e_t\}$, and then train on randomly drawn batches of samples from the memory. The memory have a finite capacity, and new experiences replace older ones when the memory is full. There are three main advantages of training on the replay memory: It is data efficient since a single experience can be drawn many times. Only training on consecutive experiences is inefficient, since the network tends to forget previous experiences by overwriting them with new experiences. The time-correlation of consecutive experiences means that the network update due to the current experience determines what the next experience will be, so training can be dominated by experiences from a certain area in the state-action space.

Target Network

Finally, we see that in Eq. 5.13 the current weights of the network determines both the target Q-value and the predicted Q-value. Thus every network update changes the target Q-value that we are trying to reach. This is like a dog chasing its own tail, and makes it hard for the network weights to converge. A simple way to

circumvent this problem is to use two neural networks, one for the target Q-value (θ^-), and one for the current Q-value (θ).

$$L(\theta) = E\left[\left(r_t + \gamma \max_{a_{t+1}} Q^*_\pi(s_{t+1}, a_{t+1}, \theta^-) - Q(s_t, a_t, \theta)\right)^2\right]. \tag{5.14}$$

The target network weights θ^- are updated to the current network weights, $\theta^- \rightarrow \theta$, every N iteration of the algorithm. A pseudocode of the DQL algorithm presented in [59] is shown below

Initialize memory M
Initialize Q-value network with weights θ
Initialize target Q-value network with weights $\theta^- = \theta$
for *episode = 1 to* N_e **do**
  Reset environment to initial state s_0
  for *t = 1 to T* **do**
    Select random action a_t with probability ϵ
    Else select $a_t = \underset{a_t}{\text{argmax}}\, Q^*(s_t, a_t)$
    Execute action in environment and observe r_t and s_{t+1}
    Store experience (s_t, a_t, r_t, s_{t+1}) in memory M
    Sample random experiences (s_i, a_i, r_i, s_{i+1}) from M
    If s_{i+1} is a terminal state set $y_i = r_i$
    Else set $y_i = r_i + \gamma \max_{a_{i+1}} Q(s_{i+1}, a_{i+1}, \theta^-)$
    Perform a gradient decent step on $(y_i - Q(s_i, a_i, \theta))^2$
    Every C steps set $\theta^- = \theta$
    Set $s_{t+1} = s_t$ and decrease ϵ
  end
end

Chapter 6

Fluctuations of biomolecular motors

In this chapter, we give a review of work done in collaboration with the Small Biosystems Lab at the University of Barcelona. The work revolves around doing experiments on the interaction between helicases and DNA. Helicases are a class of biomolecular motors proteins whose main function is to unzip the genes of an organism, i.e. its DNA or RNA [60, 61]. The one we studied belongs to the RecQ family of helicases, which unzips double-stranded DNA (dsDNA) into two complementary single-stranded DNA (ssDNA).

6.1 DNA-helicase interaction

An ssDNA consists of nucleobases (i.e., A, T, C, G) attached to a backbone of alternating phosphate and sugar residues [62]. The backbone has a directionality due to the orientation of carbon atoms in the sugar residues, and the two directions are known as the 3'-5' and the 5'-3' direction. Two ssDNA can combine to form a dsDNA (or a ssDNA can close on itself and form a hairpin loop as in Fig. 6.1), if the nucleobases of the two strands are complementary. The base-pairing rules are that adenine (A) can combine with thymine (T), and cytosine (C) can combine with guanine (G), to form nucleotide base pairs. The backbone of the two strands of a dsDNA is always anti-aligned, with respect to the 3'-5' and 5'-3' direction. When a helicase unzips a dsDNA, other bimolecular motors can attach to the single strands in order to perform various tasks. An example is the polymerase enzyme, which binds to the ssDNA and makes copies of a sequence of nucleobases. The cooperation between the helicase and polymerase is essential for DNA replication, and thus all of life as we know it.

We consider a system consisting of a 480 base pair (measured from the handles to the start of the hairpin loop) DNA-hairpin in an optical trap, which is unzipped by a RecQ-helicase, as shown in Fig. 6.1(A). Two plastic beads are attached to the ends of the hairpin, where the smaller bead is kept fixed by a suction force from a micro-pipette, and the larger bead is trapped in the focus point of an optical tweezer. We can move the bead in the optical trap, by moving the focus point, along the axis between the two beads. The bead in the trap pulls on the hairpin with a constant force feedback while the helicase translocates along it, in the 3'-5' direction, opening the base pair bonds as it goes. The helicase is powered by ATP-hydrolysis and the chemical energy supplied is

$$E_t^{ATP} = (n(t) - n(0))\, N_{ATP}\, \Delta\mu, \qquad \Delta\mu = \mu_{ATP} - \mu_{ADP} - \mu_{P_i}. \tag{6.1}$$

Here N_{ATP} is the number of ATP hydrolyzed per base pair, $n(t)$ the number of base pairs opened in time t, and $\Delta\mu$ is the change of chemical potential when converting one ATP to an ADP and an orthophosphate. The optical trap exerts a constant force feedback on the hairpin which results in the work contribution

$$W_t^{OT} = f_{OT}\,(x(t) - x(0)), \tag{6.2}$$

where f_{OT} is the constant force feedback that we apply with the optical tweezers [63]. The force applied by the optical tweezers is not large enough to mechanically break the base pair bonds. The helicase itself breaks the bonds and translocates along the hairpin. However, the force applied by the optical trap can lower the energy barrier that the helicase has to overcome to open the bonds. After a bond is broken, the force feedback quickly stretches the released extension so as to maintain a constant force on the hairpin. The feedback protocol quickly moves the trap further away to regain the constant tension force on the hairpin. The change in energy of the hairpin due to the bond breaking and strand stretching is

$$E_t^{DNA} = E_t^{bond} + E_t^{stretch} = G_0(n(t) - n(0)) - \int_0^{x(t)} f_{wlc}(x)dx, \tag{6.3}$$

where G_0 is the free energy change due to opening one bond and the integral is the work required to stretch the ssDNA handles after opening base pairs. The

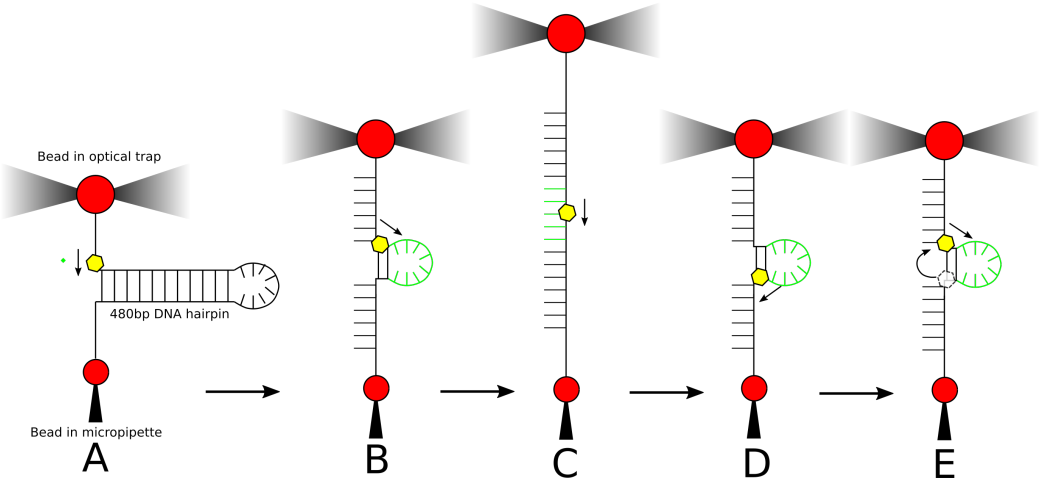

Figure 6.1: Schematic of the experimental setup. (A) shows the initial condition, just after the helicase (yellow hexagon) has attached itself to the DNA hairpin. In (B) the helicase has opened a number of base pairs, and the tweezers have stretched the released ssDNA handles. When the hairpin is fully opened (C) the helicase in the center of the loop prevents recombination of the hairpin. Once the helicase has passed the loop (D) the hairpin starts to close behind it. Finally, the helicase switches over to the complementary strand (E) and starts unzipping again.

conservation of energy therefore gives us the following relation

$$E_t^{ATP} + W^{OT} = E_t^{DNA} + Q \tag{6.4}$$

We had two main objectives to achieve in our study; to study the kinematics of the helicase (its average velocity, variance, etc.), and its energetics. We begin with the former.

6.2 Kinematics

A schematic of the hairpin-helicase interaction is shown in Fig. 6.1, and a typical experimental helicase unwinding events we observed is shown in Fig. 6.2. In Fig. 6.2 the vertical axis shows the position of the optical trap λ, where zero is the relaxed position with a fully closed hairpin. When the helicase attaches to the DNA and starts to unzip the hairpin, the optical tweezers and the bead in the trap is moved away from the bead in the micro-pipette in order to keep a constant tension

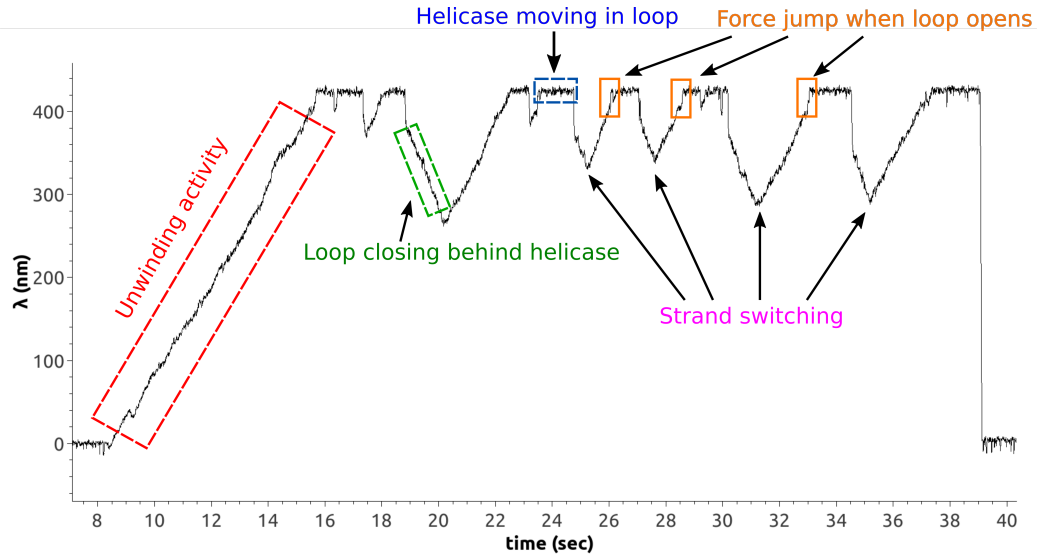

Figure 6.2: Typical event measured. In this figure we see what appears to be a single helicase, opening the hairpin, which closes behind it after it has moved past the loop, and then switching strand to open it again, for several cycles.

force on the DNA. This is indicated by the red box in Fig. 6.2, and corresponds to Fig. 6.1(A-B). When the hairpin is fully unzipped, we have a fully unwound ssDNA. The helicase then moves through the loop, blocking any recombination of the hairpin, as shown in Fig. 6.1(C) and the blue boxes in Fig. 6.2. Once the helicase has passed the loop, the hairpin starts to close again behind it (Fig. 6.1(D), and green box in Fig. 6.2). After some time, the helicase jumps over to the other strand, and begins to unzip the nucleotide base pairs again (Fig. 6.1(C)). This process of unzipping, recombination, and strand switching is repeated until the helicase disassociates from the DNA-hairpin.

We focus on the forward unzipping of the DNA, shown in the red box in Fig. 6.2, and all further analysis is performed on data of this type. The probability density of the distance moved by the helicase in a given time $\Delta t \in [25\ ms \rightarrow 800\ ms]$, when applying a constant force of $10\ pN$ is shown in Fig. 6.3. The probability densities are Gaussian for all timescales, with an increasing standard deviation as a function of time.

The helicase translocates along the DNA-hairpin, hydrolyzing ATP and opening nucleotide base pairs, and it does so at a certain average velocity. We can find this average velocity by measuring the extension of the whole hairpin with the

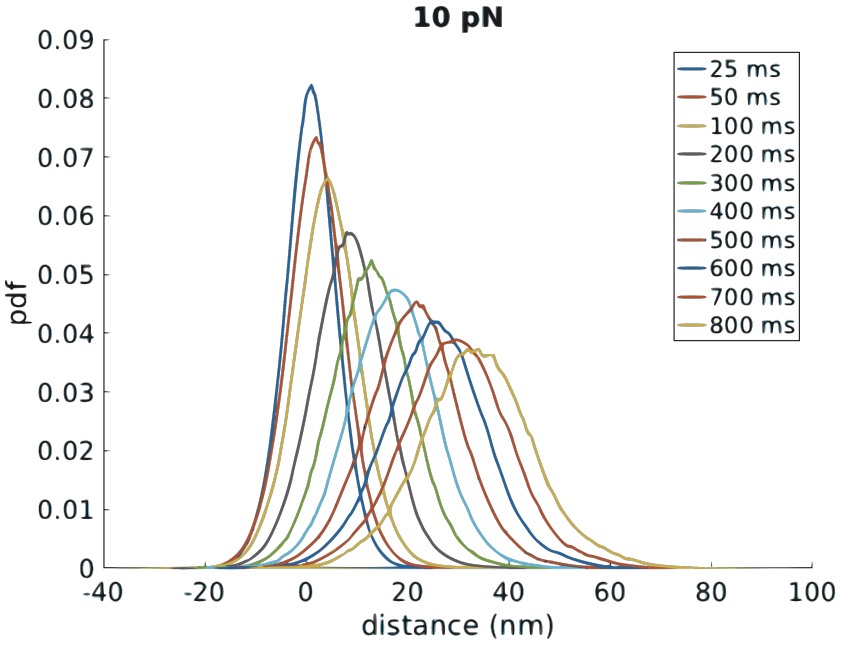

Figure 6.3: The probability distributions of the helicase moving a distance dx in a given time interval dt when a force of 10 pN is applied by the optical tweezers. The error bars are the standard deviations of the velocity for each force, and the number of unzipping events per force varies from 42 for $5\ pN$ to 146 for $9\ pN$.

optical tweezers. This measurement in units of nm/s can be used to estimate the number of base pairs that is opened per second bp/s, by using the worm-like chain model (to be introduced in section 6.3.1). Therefore we can define two different velocities; $v_x(t) = \dot{x}(t)$, which measures the rate of change of the extension of the DNA hairpin in nm/s, and $v_{bp} = \dot{n}(t)$, which measures the helicase velocity relative to the DNA substrate in units of bp/s. In Fig. 6.4 we show plots of the velocity as a function of the constant force applied by the optical tweezers.

While the rate of change for the extension (measured in nm/s) is an increasing function of force, the velocity of the helicase with respect to the DNA substrate (measured in bp/s) is approximately independent of force. For high forces the end-to-end extension of the ssDNA handles (which is what we measure with the optical tweezers), is approximately the same as the contour length, i.e. the ssDNA is fully stretched. Therefore any change in the extension we measure is due to the opening of base pairs. However, when the force is decreased, such that it is lower than the entropic force pushing the polymer towards a more entangled configuration, the ssDNA forms bends and curls. Therefore, when the helicase breaks base pair bonds in the low force regime, the full length of the newly released segment does not result in the same length change of the end-to-end distance, but rather allows for more twists and bends in the ssDNA. This is the reason why the velocity of the end-to-end extension $v_x(t)$ decreases for lower forces.

6.3 Energetics

6.3.1 Work cost of stretching DNA polymer

At finite temperatures, the distance between the two ends of a ssDNA (end-to-end distance) will be significantly shorter than the contour length L_0. This is caused by thermal fluctuations, which result in a coiled, random configuration of the polymer. Upon stretching the polymer, the accessible spectrum of fluctuations reduces, which causes an entropic force against the external elongation. The worm-like chain (WLC) model is an interpolation formula that approximates the force-extension behavior, and is given by

$$fP\beta = \frac{1}{4}\left(1 - \frac{x}{L_0}\right)^{-2} - \frac{1}{4} + \frac{x}{L_0}, \tag{6.5}$$

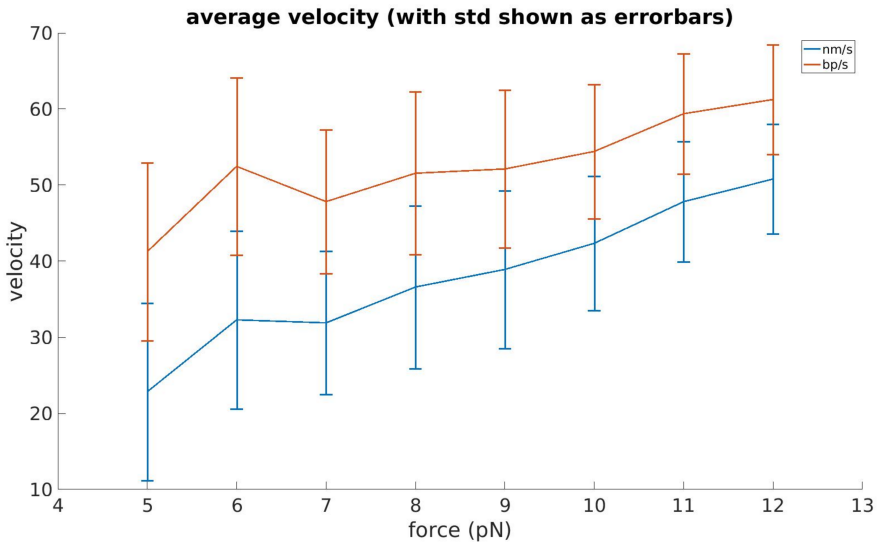

Figure 6.4: Velocity as a function of force. The orange line is in units of bp/s (the actual velocity of the helicase), and blue line is in units of nm/s (the rate of change of the extension of the DNA hairpin as measured by the optical tweezers).

where x is the end-to-end distance, and P the persistence length, β is the inverse temperature, and f is the external force applied. The persistence length is the length at which correlations in the direction of the tangent of the polymer are lost. The ratio x/L_0 can be considered as the fraction of "stretched" polymer. If it is equal to 1, there are no coils and the entropy is zero.

Finding the equilibrium extension at a given force

To find the end-to-end distance of a polymer with contour length L_0, and an external force f, we rewrite Eq. (6.5) as a third order polynomial and relabel the variable $x' = x/L_0$

$$(1 + 4fP\beta)\left(1 - x'\right)^2 - 4x'\left(1 - x'\right)^2 - 1 = 0 \tag{6.6}$$

$$(1 + 4fP\beta)\left(x'^2 - 2x' + 1\right) - 4x'\left(x'^2 - 2x' + 1\right)^2 - 1 = 0 \tag{6.7}$$

relabeling $C = 1 + 4fP\beta$, and gathering the terms we get

$$4x'^3 - (C + 8)x'^2 - (2C + 4)x' + (1 - C) = 0, \tag{6.8}$$

which is a third order polynomial, which has one real and positive solution. If the polymer is a single-stranded DNA, we can estimate the contour length as $L_0 = N_b\ l_{bp}$, where N_b is the number of nucleotide monomers in the strand, and l_{bp} is the length of one nucleotide. One of the roots of Eq. (6.8) is some real positive number X_0, which we can use to estimate the number of bases in a ssDNA. With the root $X_0 = x/L_0$ we find

$$\frac{x}{L_0} = \frac{x}{N_b\ l_{bp}} = X_0 \qquad \rightarrow \qquad N_b = \frac{x}{X_0\ l_{bp}}, \tag{6.9}$$

where x is the extension at a given force, found from experimental measurements.

To summarize, the worm-like change model is a relationship between the external force, end-to-end distance, and contour length of a polymer; if you know two of the variables you can find the third. The minimum work done by stretching the polymer from an end-to-end distance x_0 to x_t can be found by integrating Eq. (6.5)

$$W_s = \int_{x_0}^{x_t} f\ dx = \frac{1}{P\beta}\left[\frac{L_0^2}{4(L_0 - x)} - \frac{x}{4} + \frac{x^2}{2\ L_0}\right]_{x_0}^{x_t}. \tag{6.10}$$

which in the case where $x_0 = 0$ becomes

$$W_s = \frac{1}{P\beta}\left[\frac{L_0^2}{4(L_0 - x_t)} - \frac{(x_t + L_0)}{4} + \frac{x_t^2}{2\ L_0}\right] \tag{6.11}$$

Stretching of a DNA-hairpin

In the previous section, the assumption was that the polymer is static with constant contour length. When we consider a DNA-hairpin the interpretation changes a bit. In that case the contour length is a function of time $L_0(t) = N_{bp}(t)\ l_{bp}$, because the number of nucleotides opened depends in some way on time (either due to time dependence in the force applied, or due to helicase unwinding activity). However, assuming that the hairpin itself does not significantly affect the coiling behavior of the ssDNA handles, we can still use the WLC model. The math would be the same as in the previous section, up until we find X_0. Now every time a base pair is opened, we gain a length that is equal to two nucleotides, therefore the ratio between a change in extension Δx and the associated change in number of base pairs opened ΔN_{bp} is

$$\frac{\Delta x}{\Delta L_0} = \frac{\Delta x}{2 l_{bp} \Delta N_{bp}} = X_0, \tag{6.12}$$

We can define a constant γ, given by

$$\gamma = \frac{\Delta x}{\Delta N_{bp}} = 2X_0\, l_{bp} \tag{6.13}$$

which can be considered as a conversion factor between the end-to-end distance Δx and the number of base pairs opened in the hairpin, ΔN_{bp}. This factor can be used to find how many base pairs are opened when the extension increases by an amount Δx. Assume now we have measured an increase in the end-to-end distance Δx, and we want to find the minimum amount of work that was performed in stretching that new segment. From Eq. (6.11) we find

$$\begin{aligned} W_s &= \frac{1}{P\beta}\left[\frac{L_0^2}{4(L_0-\Delta x)} - \frac{(\Delta x + L_0)}{4} + \frac{\Delta x^2}{2\,L_0}\right] \\ &= \frac{1}{P\beta}\left[\frac{(2\Delta N_{bp} l_{bp})^2}{4(2\Delta N_{bp} l_{bp} - \Delta x)} - \frac{(\Delta x + 2\Delta N_{bp} l_{bp})}{4} + \frac{\Delta x^2}{2\,(2\Delta N_{bp} l_{bp})}\right] \\ &= \frac{\Delta x}{P\beta\gamma}\left[\frac{l_{bp}^2}{2l_{bp}-\gamma} - \frac{(\gamma + 2l_{bp})}{4} + \frac{\gamma^2}{4l_{bp}}\right] \end{aligned} \tag{6.14}$$

where we have used $\Delta x = \Delta N_{bp}\gamma$ and $\Delta L_0 = 2\Delta N_{bp} l_{bp}$. The net work done is therefore

$$W_{net} = W_{OT} - W_S = \Delta x\left[f - \frac{1}{P\beta\gamma}\left(\frac{l_{bp}^2}{2l_{bp}-\gamma} - \frac{(\gamma+2l_{bp})}{4} + \frac{\gamma^2}{4l_{bp}}\right)\right] \tag{6.15}$$

6.3.2 Entropy production and fluctuations

Steady state entropy production

In the introduction of this chapter, we briefly discussed the energetics of the combined helicase and dsDNA system. According to the first law of thermodynamics, the energy balance is

$$\frac{\Delta x\; N_{ATP}\; \Delta\mu}{\gamma} + f\Delta x = \Delta G + Q, \tag{6.16}$$

where ΔG is the change of free energy of the dsDNA obtained when the helicase unzips Δbp base pairs, or equivalently the bead in the optical tweezer moves a distance $\Delta x = \gamma \Delta bp$

$$\Delta G = \Delta G_0 + \int_{x_0}^{x_t} f_{wlc}\, dx. \tag{6.17}$$

Here $\Delta G_0 = \frac{\Delta x \Delta G_{bp}}{\gamma}$ is the change in free energy due to base pair bond breaking (ΔG_{bp} is the change in free energy per bp, $\sim 2k_BT$ at room temperature), while the integral is the stretching contribution. Solving this for Q and dividing by T we get the entropy production as

$$\Delta S = \frac{1}{T}\left[\frac{\Delta x\ N_{ATP}\ \Delta\mu}{\gamma} - \frac{\Delta x \Delta G_{bp}}{\gamma} + f\Delta x - \int_{x_0}^{x_t} f_{wlc}\, dx.\right] \quad (6.18)$$

$$= \underbrace{\frac{\Delta x}{T\gamma}\left(N_{ATP}\Delta\mu - \Delta G_{bp}\right)}_{S_{bp}} + \underbrace{\frac{1}{T}\left(f\Delta x - \int_{x_0}^{x_t} f_{wlc}\, dx\right)}_{S_{stretch}} \quad (6.19)$$

The equation is divided into two parts, which helps us think about the different contributions to the total entropy production.

S_{bp} is the entropy production associated with the inefficiency of the helicase motor. In the reversible case the energy gained by ATP hydrolysis is 100% used to open base pairs so the term in the parenthesis becomes zero

$$N_{ATP}\Delta\mu - \Delta G_{bp} = 0,$$

meaning no heat is dissipated in the process and hence no entropy is produced. In the case where

$$N_{ATP}\Delta\mu - \Delta G_{bp} > 0,$$

we have a contribution to the total entropy production. This contribution is always positive, or else it would violate the second law of thermodynamics.

$S_{stretch}$ is the entropy production associated with the work performed to stretch the DNA after base pairs have opened. In the reversible case where

$$f\Delta x - \int_{x_0}^{x_t} f_{wlc}\, dx = 0$$

we use an amount of work $f\Delta x = f(x_t - x_0)$, to increase the free energy of the DNA by an amount $\int_{x_0}^{x_t} f_{wlc}\, dx$, but no heat (or entropy) is generated. All of the work is reversibly converted into potential energy. Now, if we have

$$f\Delta x_t - \int_{x_0}^{x_t} f_{wlc}\, dx > 0,$$

heat is dissipated since we apply more force than needed to stretch the DNA. Again this contribution is always positive due to the second law.

The integral in the stretching term was found in section 6.3.1. Putting this result into the entropy formula we get

$$\Delta S_t = \frac{\Delta x}{T}\left[\frac{N_{ATP}\Delta\mu - \Delta G_{bp}}{\gamma} + f - \frac{1}{\beta P}\left(\frac{l_{bp}^2}{4\gamma(l_{bp}-\gamma)} - \frac{(1+\frac{l_{bp}}{\gamma})}{4} + \frac{\gamma}{2l_{bp}}\right)\right] \tag{6.20}$$

We now have an entropy production that is linearly dependent on Δx,

$$S = A\Delta x,$$

where the slope A is given by

$$A = \frac{1}{T}\left[\frac{N_{ATP}\Delta\mu - \Delta G_{bp}}{\gamma} + f - \frac{1}{\beta P}\left(\frac{l_{bp}^2}{4\gamma(l_{bp}-\gamma)} - \frac{(1+\frac{l_{bp}}{\gamma})}{4} + \frac{\gamma}{2l_{bp}}\right)\right] \tag{6.21}$$

Fluctuation theorem

We now assume that the entropy production obeys the standard fluctuation theorem [15]

$$\frac{P(\Delta S)}{P(-\Delta S)} = \exp\left[\Delta S/k_B\right], \tag{6.22}$$

where $P(\Delta S)$ is the probability to generate an amount of entropy ΔS, while $P(-\Delta S)$ is the probability to consume an amount of entropy $-\Delta S$. In this formulation ΔS only varies in time due to variations in the extension Δx, which again depends on time. All other variables are constant in time. The ratio of the probability distributions of ΔS and Δx are therefore identical, and we have

$$\frac{P(\Delta S)}{P(-\Delta S)} = \frac{P(\Delta x)}{P(-\Delta x)} = e^{\Delta S/k_B} = e^{A\Delta x/k_B}, \tag{6.23}$$

where A is defined in Eq. (6.21). For non-equilibrium steady states (NESS), the fluctuation theorem takes the form [64]

$$\lim_{t\to\infty} \underbrace{\frac{1}{\langle\Delta S\rangle}\ln\left(\frac{P(a)}{P(-a)}\right)}_{f_t(a)} = a, \tag{6.24}$$

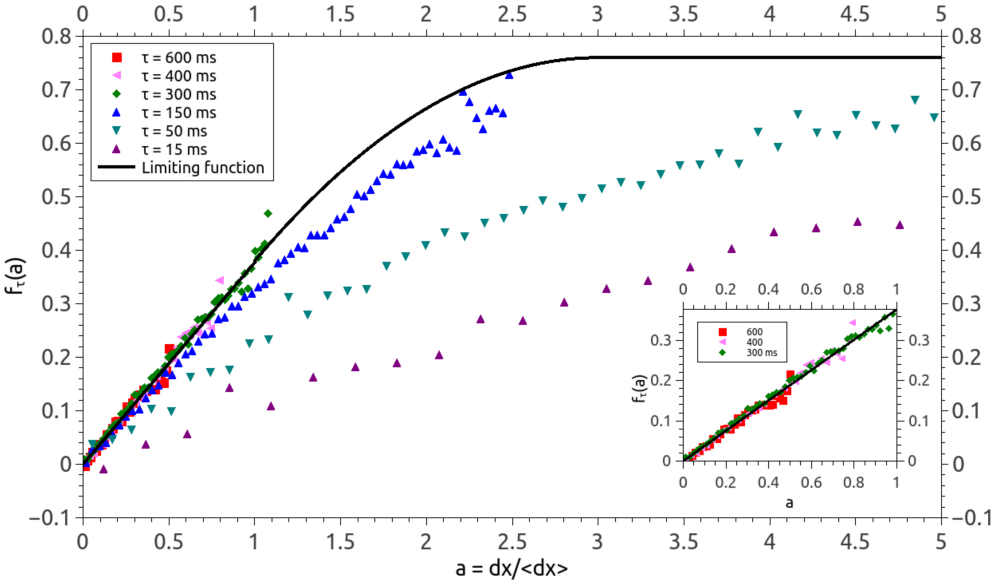

Figure 6.5: Plot of $f_t(a)$ as a function of a, for increasing time intervals from $\tau \in [15\ ms,\ 600\ ms]$. As the time interval increases, the curves converges to a linear function of time with fixed slope, as predicted by the non-equilibrium steady state fluctuation theorem.

where $a = \Delta S/\langle \Delta S \rangle$. By taking the logarithm of Eq. (6.23) and dividing by $\langle \Delta x \rangle$, we obtain the NESS fluctuation theorem for our system

$$f_t(a)\frac{1}{\langle \Delta x \rangle} \ln\left[\frac{P(a)}{P(-a)}\right] = a\frac{A}{k_B}, \tag{6.25}$$

where a is now $a = \Delta x/\langle \Delta x \rangle$. If we plot the left-hand side of Eq. (6.25) as a function of a, we see that we can read of the value of constant A from the slope of a linear fit. In Fig. 6.5 we plot $f_t(a)$ for different time-intervals τ, when we apply a constant feedback force of $12\ pN$. For larger time-intervals, $f_t(a)$ converges to a linear function, with a fixed slope. The slope of $f_t(a)$ converges to a fixed value as $t \to \infty$, as shown in Fig. 6.6.

Moreover, with the experimentally obtained value for the constant A, we can use Eq. (6.21) to obtain the number of ATP hydrolyzed per base pair opened by

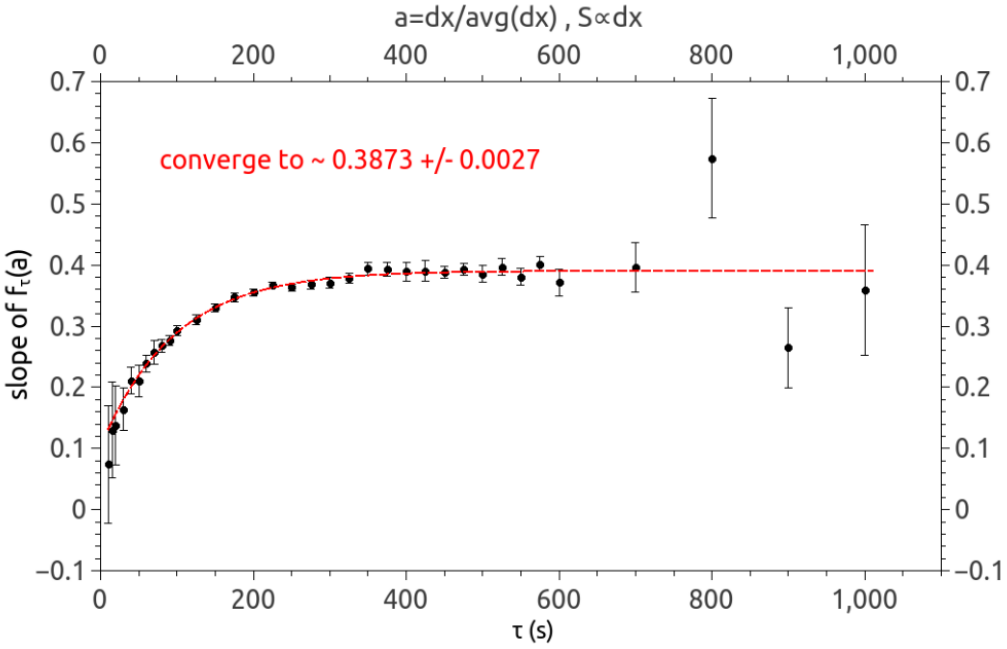

Figure 6.6: Plot of the slope of a linear fit of $f_t(a)$, as a function of the time interval τ. As the time interval increases, the slope of the linear fit converges to a fixed value, as predicted by the non-equilibrium steady state fluctuation theorem.

the helicase

$$N_{ATP} = \frac{\gamma}{\Delta\mu}\left[A\,kT - f + \frac{1}{\beta P}\left(\frac{l_{bp}^2}{4\gamma(l_{bp}-\gamma)} - \frac{(1+\frac{l_{bp}}{\gamma})}{4} + \frac{\gamma}{2l_{bp}}\right)\right] + \frac{\Delta G_{bp}}{\Delta\mu}. \tag{6.26}$$

Using this equation for the data obtained from $12\ pN$ constant feedback, we find that the number of base pairs opened per ATP hydrolyzed is about $\sim 14\ bp/ATP$. The hydrolyzation of one ATP provides about $\sim 20\ k_BT$ worth of energy, while it costs on average $\sim 2\ k_BT$ to break a nucleotide bond. This implies that if the helicase is operating at 100% efficiency, it can open a maximum of ~ 10 base pairs per ATP. Since the number we get is higher than this value, it implies that the tension applied to the nucleotide bonds by the constant force feedback from the optical tweezers reduces the energetic cost of breaking the bonds.

The problem with the results we obtain by using Eq. (6.20), is that we assume the only thing the helicase does when it hydrolyzes ATP is that it translocates. However, the helicase could be using the ATP for conformational changes of the molecule. Typically, the operation of molecular motors requires successive conformal changes, like kinesin "walking" on a microtubule. To account for this source of entropy production we have to add another unknown "hidden" entropy production term, that is assumed to be linear in time.

$$\Delta S_t = \underbrace{\frac{\Delta x_t}{T\gamma}\left(N_{ATP}\Delta\mu - \Delta G_{bp}\right)}_{S_{bp}} + \underbrace{\frac{1}{T}\left(f\Delta x_t - \int_{x_0}^{x_t} f_{wlc}\,dx\right)}_{S_{stretch}} + \underbrace{D\Delta t}_{S_{hidden}} \tag{6.27}$$

Here D is a constant dissipation rate, which we associate with ATP hydrolysis which does not result in translation of the helicase. However, since the fluctuation in the number of base pairs hydrolyzed is not entirely due to the fluctuation in the change in extension of the DNA hairpin, we can no longer use the fluctuation theorem on the extension data, and we will have to find an alternate method for analyzing the data.

Appendix A

Energy dependent tunneling rate

In our paper on the optimal protocol (paper 2) for a Szilard engine with measurement errors, we assumed the tunneling rate, Γ, was constant. The experiment we based our analytical work on [41], is a single-electron box, consisting of two metallic islands connected by a metal/insulator/superconductor (NIS) tunnel junction. The tunneling rate of the NIS junction is [65]

$$\Gamma(V) = 2\Gamma_m \cosh^2(V/2), \qquad \Gamma_m = \frac{\sqrt{2\pi\Delta k_B T}}{R_T e^2} e^{-\Delta/k_B T}, \tag{A.1}$$

where R_T is the normal-state resistance of the NIS tunnel junction and Δ is the superconducting energy gap in the superconducting electrode of the junction. Here we update our model by using this V dependent tunneling rate, $\Gamma[V(t)]$. We include all the details of the necessary calculations, which can also be used to reproduce our results for the constant Γ case. Details that are already in the paper is left out, so in order to follow the coming calculations, we recommend reading the paper first.

A.1 Deriving the differential equation

Beginning from the master equation

$$\begin{aligned} \dot{p}_1 &= -\Gamma_{12}p_1 + \Gamma_{21}p_2 = -\Gamma p_1 + \Gamma_{21}, \\ \dot{p}_2 &= \Gamma_{12}p_1 - \Gamma_{21}p_2 = -\Gamma p_2 + \Gamma_{12}, \end{aligned} \tag{A.2}$$

where $\Gamma = \Gamma_{12} + \Gamma_{21}$. Detailed balance gives us

$$\frac{\Gamma_{21}}{\Gamma_{12}} = e^V \quad \rightarrow \quad \Gamma_{12} = (1+e^V)^{-1}\Gamma \tag{A.3}$$

Putting this in Eq. (A.3) for $\dot{p}_2$ we get

$$\dot{p}_2 = \Gamma(t)\left[\frac{1}{1+e^V} - p_2\right] \tag{A.4}$$

using the time NIS tunneling rate, this becomes

$$\dot{p}_2 \equiv \dot{p} = 2\Gamma_m \cosh^2(V/2)\left[\frac{1}{1+e^V} - p\right] \tag{A.5}$$

Implementing this into our existing problem is easy. We only have to change our expression for $\dot{p}$, which we update for each step when solving the differential equation.

It is more difficult when inverting this equation to solve V as a function of p and $\dot{p}$, which we need if we want to solve the problem in terms of p and $\dot{p}$ as we did previously. Rewriting we obtain

$$\begin{aligned}\dot{p} &= 2\Gamma_m \cosh^2(V/2)\left[\frac{1}{1+e^V} - p\right] \\ &= \Gamma_m\left[e^{-V/2}\cosh(V/2) - p\left(\cosh(V)+1\right)\right]\end{aligned} \tag{A.6}$$

The power is still given by

$$P = \frac{1}{\tau}\int_0^\tau dt\, \dot{p}V \tag{A.7}$$

but now we can not write V purely as a function of p and $\dot{p}$. We have to change variables as compared to how we do it for the constant Γ case. There we had p and $\dot{p}$ as the dependent variables of the Lagrangian, but now we have to use either (p, V) or $(\dot{p}, V)$, and redo all the calculations. Using (p, V), we get

$$P = \frac{1}{\tau}\int_0^\tau dt\, \dot{p}V \tag{A.8}$$

$$= \frac{2}{\tau}\int_0^\tau dt\, V\Gamma_m \cosh^2(V/2)\left[\frac{1}{1+e^V} - p\right] \tag{A.9}$$

$$= \frac{\Gamma_m}{\tau}\int_0^\tau dt\, V\left(\frac{1}{2}(e^{-V}+1) - p\left[\cosh(V)+1\right]\right) \tag{A.10}$$

for the power, while for the change in information entropy we get

$$\begin{aligned}\frac{\Delta H}{\tau} &= -\frac{1}{\tau}\int_0^\tau dt\,\dot{p}\ln\left(\frac{p}{1-p}\right) \\ &= -\frac{2}{\tau}\int_0^\tau dt\ln\left(\frac{p}{1-p}\right)\Gamma_m\cosh^2\left(V/2\right)\left[\frac{1}{1+e^V}-p\right] \\ &= -\frac{\Gamma_m}{\tau}\int_0^\tau dt\ln\left(\frac{p}{1-p}\right)\left(\frac{1}{2}(e^{-V}+1)-p\left[\cosh(V)+1\right]\right)\end{aligned} \tag{A.11}$$

The functional we want to minimize therefore becomes

$$J = \frac{S_\epsilon}{\tau}+\frac{\Delta H}{\tau}+\lambda P = \frac{S_\epsilon}{\tau}+\frac{1}{\tau}\int_0^\tau dt\;L(p,\dot{p},\lambda), \tag{A.12}$$

with the Lagrangian

$$L(V,p,\dot{p},\lambda) = \dot{p}\left[-\ln\left(\frac{p}{1-p}\right)+\lambda V\right], \tag{A.13}$$

which can be written in terms of (p,V) as

$$\begin{aligned}L(V,p,,\lambda) &= -2\Gamma_m\ln\left(\frac{p}{1-p}\right)\cosh^2\left(V/2\right)\left[\frac{1}{1+e^V}-p\right] \\ &+ 2\lambda V\Gamma_m\cosh^2\left(V/2\right)\left[\frac{1}{1+e^V}-p\right] \\ &= 2\Gamma_m\cosh^2\left(V/2\right)\left[\frac{1}{1+e^V}-p\right]\left[\lambda V-\ln\left(\frac{p}{1-p}\right)\right].\end{aligned} \tag{A.14}$$

Using the Euler-Lagrange formalism

$$\frac{\partial L}{\partial p} = \frac{d}{dt}\frac{\partial L}{\partial \dot{p}} \tag{A.15}$$

on Eq. (A.13), the left-hand side becomes

$$\frac{\partial L}{\partial p} = \dot{p}\left[\lambda\frac{\partial V}{\partial p}-\frac{1}{p(1-p)}\right] = \frac{\dot{p}}{p(p-1)}+\lambda\dot{p}\frac{\partial V}{\partial p} \tag{A.16}$$

We have

$$\dot{p} = 2\Gamma_m\cosh^2\left(V/2\right)\left[\frac{1}{1+e^V}-p\right], \tag{A.17}$$

and taking by the partial derivative with respect to p here we get

$$\frac{\partial}{\partial p}\left[2\Gamma_m \cosh^2(V/2)\left[\frac{1}{1+e^V}-p\right]\right]=0$$
$$2\Gamma_m \cosh^2(V/2)\left[-\frac{\partial V}{\partial p}\frac{e^V}{(e^V+1)^2}-1\right]+\frac{\partial V}{\partial p}\Gamma_m \sinh(V)\left[\frac{1}{1+e^V}-p\right]=0$$
$$\frac{\partial V}{\partial p}\left[\frac{2\cosh^2(V/2)e^V}{(e^V+1)^2}-\sinh(V)\left[\frac{1}{1+e^V}-p\right]\right]=-2\cosh^2(V/2)$$

Solving for $\partial V/\partial p$, we obtain

$$\begin{aligned}\frac{\partial V}{\partial p} &= \frac{2\cosh^2(V/2)}{\sinh(V)\left[\frac{1}{1+e^V}-p\right]-\frac{2\cosh^2(V/2)e^V}{(e^V+1)^2}} \\ &= \frac{2\cosh^2(V/2)}{\sinh(V)\left[\frac{1}{1+e^V}-p\right]-\frac{1}{2}}\end{aligned} \tag{A.18}$$

The right-hand side of the Euler-Lagrange equation becomes

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{p}}=\frac{d}{dt}\left[-\ln\left(\frac{p}{1-p}\right)+\lambda V+\lambda\dot{p}\frac{\partial V}{\partial \dot{p}}\right]. \tag{A.19}$$

We can find $\partial V/\partial \dot{p}$ by using a property of partial derivatives

$$\frac{\partial V}{\partial \dot{p}}=\frac{\partial V}{\partial p}\frac{\partial p}{\partial \dot{p}}, \tag{A.20}$$

and since

$$\frac{\partial p}{\partial \dot{p}} = \frac{\partial}{\partial \dot{p}}\left[\frac{\dot{p}}{2\Gamma_m \cosh^2(V/2)}-\frac{1}{1+e^V}\right] \tag{A.21}$$
$$= \frac{1}{2\Gamma_m \cosh^2(V/2)}, \tag{A.22}$$

we obtain the following relationship between $\partial V \partial p$ and $\partial V/\partial \dot{p}$:

$$\frac{\partial V}{\partial p}=2\cosh^2(V/2)\frac{\partial V}{\partial \dot{p}} \tag{A.23}$$

We can now perform the time derivate of the right-hand side, and obtain

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{p}} = \frac{\dot{p}}{p(p-1)} + \lambda\left[\dot{V} + \dot{p}\frac{d}{dt}\frac{\partial V}{\partial \dot{p}} + \ddot{p}\frac{\partial V}{\partial \dot{p}}\right]. \tag{A.24}$$

Writing the full Euler-Lagrange equation in terms of the expressions we have obtained so far, we have

$$\dot{p}\frac{\partial V}{\partial p} = \frac{dV}{dt} + \dot{p}\frac{d}{dt}\left(\frac{\partial V}{\partial \dot{p}}\right) + \ddot{p}\frac{\partial V}{\partial \dot{p}}, \tag{A.25}$$

and we see that the factors with the Lagrange multiplier λ on the left and right-hand side cancel each other.
Now we need to calculate $\frac{dV}{dt}$. From Eq. (A.17) we get

$$\begin{aligned}
\ddot{p} &= \frac{d}{dt}\left[2\cosh^2(V/2)\left(\frac{1}{1+e^V} - p\right)\right] \\
&= \dot{V}\sinh(V)\left[\frac{1}{1+e^V} - p\right] - 2\cosh^2(V/2)\left[\frac{e^V\dot{V}}{(1+e^V)^2} + \dot{p}\right] \\
&= \dot{V}\sinh(V)\left[\frac{1}{1+e^V} - p\right] - \frac{\dot{V}}{2} - 2\dot{p}\cosh^2(V/2)
\end{aligned} \tag{A.26}$$

and solving for $\dot{V}$ we obtain

$$\dot{V} = \frac{\ddot{p} + 2\dot{p}\cosh^2(V/2)}{\sinh V\left[\frac{1}{1+e^V} - p\right] - \frac{1}{2}}. \tag{A.27}$$

The final thing we need to calculate is $\frac{d}{dt}\frac{\partial V}{\partial \dot{p}}$:

$$\begin{aligned}
\frac{d}{dt}\frac{\partial V}{\partial \dot{p}} &= \frac{d}{dt}\left[\frac{1}{\left(\frac{1}{1+e^V} - p\right)\sinh V - \frac{1}{2}}\right] \\
&= -\frac{\dot{V}\cosh V\left(\frac{1}{1+e^V} - p\right) + \sinh V\left(-\frac{\dot{V}e^V}{(1+e^V)^2} - \dot{p}\right)}{\left(\left(\frac{1}{1+e^V} - p\right)\sinh V - \frac{1}{2}\right)^2} \\
&= \frac{\dot{V}\left[\frac{1}{2}\tanh(V/2) - \cosh V\left(\frac{1}{1+e^V} - p\right)\right] + \dot{p}\sinh V}{\left(\left(\frac{1}{1+e^V} - p\right)\sinh V - \frac{1}{2}\right)^2}
\end{aligned} \tag{A.28}$$

Putting all the expressions we have calculated so far into Eq. (A.25) we finally

obtain the differential equation we need to solve:

$$\begin{aligned}\dot{p}\left[\frac{2\cosh^2(V/2)}{\sinh V\left[\frac{1}{1+e^V}-p\right]-\frac{1}{2}}\right] &= \frac{\ddot{p}+2\dot{p}\cosh^2(V/2)}{\sinh V\left[\frac{1}{1+e^V}-p\right]-\frac{1}{2}} \\ &+ \dot{p}\frac{\dot{V}\left[\frac{1}{2}\tanh(V/2)-\cosh V\left(\frac{1}{1+e^V}-p\right)\right]+\dot{p}\sinh V}{\left(\sinh V\left[\frac{1}{1+e^V}-p\right]-\frac{1}{2}\right)^2} \\ &+ \ddot{p}\frac{1}{\sinh V\left[\frac{1}{1+e^V}-p\right]-\frac{1}{2}} \end{aligned} \tag{A.29}$$

This is a very messy equation, but it can be simplified considerably. Multiplying by the denominator and gathering terms obtain

$$2\ddot{p}=\dot{p}\frac{\dot{V}\left[\cosh V\left(\frac{1}{1+e^V}-p\right)-\frac{1}{2}\tanh(V/2)\right]-\dot{p}\sinh V}{\sinh V\left[\frac{1}{1+e^V}-p\right]-\frac{1}{2}} \tag{A.30}$$

We now insert for $\dot{V}$

$$\begin{aligned}2\ddot{p}/\dot{p} &= \frac{\frac{\ddot{p}+2\dot{p}\cosh^2(V/2)}{\sinh V\left[\frac{1}{1+e^V}-p\right]-\frac{1}{2}}\left[\cosh V\left(\frac{1}{1+e^V}-p\right)-\frac{1}{2}\tanh(V/2)\right]-\dot{p}\sinh V}{\sinh V\left[\frac{1}{1+e^V}-p\right]-\frac{1}{2}} \\ &= \frac{\left(\ddot{p}+2\dot{p}\cosh^2(V/2)\right)\left[\frac{e^{-V}}{2}-p\cosh V\right]-\dot{p}\sinh V\left[\sinh V\left[\frac{1}{1+e^V}-p\right]-\frac{1}{2}\right]}{\left(\sinh V\left[\frac{1}{1+e^V}-p\right]-\frac{1}{2}\right)^2}\end{aligned} \tag{A.31}$$

Multiplying by the denominator and $\dot{p}$ on both sides we get

$$\begin{aligned}2\ddot{p}\left(\sinh V\left[\frac{1}{1+e^V}-p\right]-\frac{1}{2}\right)^2 &= \left(\dot{p}\ddot{p}+2\dot{p}^2\cosh^2(V/2)\right)\left[\frac{e^{-V}}{2}-p\cosh V\right] \\ &- \dot{p}^2\sinh V\left[\sinh V\left[\frac{1}{1+e^V}-p\right]-\frac{1}{2}\right]\end{aligned}$$

Gathering $\ddot{p}$ terms on the left-hand side and $\dot{p}$ terms on the right-hand side

$$\begin{aligned}
&\ddot{p}\left[2\left(\sinh V\left[\frac{1}{1+e^{V}}-p\right]-\frac{1}{2}\right)^{2}+\dot{p}\left(p\cosh V-\frac{e^{-V}}{2}\right)\right]\\
=\;&\dot{p}^{2}\left[2\cosh^{2}(V/2)\left(\frac{e^{-V}}{2}-p\cosh V\right)-\sinh V\left(\sinh V\left[\frac{1}{1+e^{V}}-p\right]-\frac{1}{2}\right)\right]\\
=\;&\dot{p}^{2}\left[\frac{1}{2}\left(1+e^{-V}\right)-p\left(1+\cosh V\right)\right]
\end{aligned}$$

And finally, by isolating $\ddot{p}$ on the left-hand side, we get the following second order differential equation

$$\ddot{p}=\dot{p}^{2}\frac{\left(1+e^{-V}\right)-2p\left(1+\cosh V\right)}{\left((2p-1)\sinh V+\cosh(V)\right)^{2}+\dot{p}\left(2p\cosh V-e^{-V}\right)} \tag{A.32}$$

A.2 Constraints

Fixed power

Since we are interested in optimal protocols with finite power, we have a constraint on the heat flux (power) from the environment. This constraint is solved numerically, similarly as in the case of constant Γ, but now it becomes

$$\begin{aligned}
G &= P-\frac{1}{\tau}\int_{0}^{\tau}dt\;\dot{p}V=0 \qquad (A.33)\\
G &= P-\frac{2}{\tau}\int_{0}^{\tau}dt\;V\Gamma_{m}\cosh^{2}\left(V/2\right)\left[\frac{1}{1+e^{V}}-p\right]=0\\
G &= P-\frac{\Gamma_{m}}{\tau}\int_{0}^{\tau}dt\;V\left(\frac{1}{2}(e^{-V}+1)-p\left[\cosh(V)+1\right]\right)=0
\end{aligned}$$

Free endpoint

The value of $p(t)$ is not fixed at the endpoint τ, therefore we have the following constraint:

$$\left(\frac{\partial L}{\partial\dot{p}}\right)_{t=\tau}=\left(\frac{\partial L}{\partial\dot{p}}\right)_{t=\tau}=-\ln\left(\frac{p_{\tau}}{1-p_{\tau}}\right)+\lambda\left[V_{\tau}+\dot{p}_{\tau}\left(\frac{\partial V}{\partial\dot{p}}\right)_{\tau}\right]=0, \tag{A.34}$$

where the subscript τ means that the variable should be evaluated at time $t = \tau$, and

$$\left(\frac{\partial V}{\partial \dot{p}}\right)_\tau = \frac{1}{\left(\frac{1}{1+e^{V_\tau}} - p_\tau\right)\sinh(V_\tau) - \frac{1}{2}}.$$

Solving for λ, this constraint becomes

$$\lambda = \frac{\ln\left(\frac{p_\tau}{1-p_\tau}\right)}{V_\tau + \dot{p}_\tau\left(\left[\frac{1}{1+e^{V_\tau}} - p_\tau\right]\sinh(V_\tau) - \frac{1}{2}\right)^{-1}} \tag{A.35}$$

Variation of J with respect to τ

The final constraint is due to the fact that the variation of J with respect to the period τ should be zero. It is given by

$$\frac{\partial J}{\partial \tau} = \lambda\frac{\partial P}{\partial \tau} + \frac{\partial \Delta H}{\partial \tau} - \frac{S_\epsilon}{\tau^2} = 0 \tag{A.36}$$

Here $\frac{\partial \Delta H}{\partial \tau}$ is the same as with constant gamma

$$\frac{\partial \Delta H}{\partial \tau} = \frac{\partial}{\partial \tau}\left[-\frac{1}{\tau}\int_0^\tau dt\dot{p}\ln\left(\frac{p}{1-p}\right)\right] = -\frac{1}{\tau}\dot{p}_\tau\ln\left(\frac{p_\tau}{1-p_\tau}\right) - \frac{\Delta H}{\tau^2}$$

but for $\frac{\partial P}{\partial \tau}$ we get

$$\frac{\partial P}{\partial \tau} = \frac{\partial}{\partial \tau}\left[\frac{1}{\tau}\int_0^\tau dt\dot{p}V\right] = -\frac{P}{\tau} + \frac{\dot{p}_\tau V_\tau}{\tau}$$

Expressed in terms of these variables, the constraint becomes

$$\lambda\left[-\frac{P}{\tau} + \frac{\dot{p}_\tau V_\tau}{\tau}\right] - \frac{1}{\tau}\dot{p}_\tau\ln\left(\frac{p_\tau}{1-p_\tau}\right) - \frac{[\Delta H + S_\epsilon]}{\tau^2} = 0$$

$$\lambda = \frac{\dot{p}_\tau\ln\left(\frac{p_\tau}{1-p_\tau}\right) + \frac{[\Delta H + S_\epsilon]}{\tau}}{[\dot{p}_\tau V_\tau - P]} \tag{A.37}$$

Combining the free endpoint and variation of τ constraints

We can combine the free endpoint and $\partial J/\partial\tau$ constraints by canceling the Lagrange multiplier λ which is found in both constraints, in the following way

$$\frac{\ln\left(\frac{p_\tau}{1-p_\tau}\right)}{V_\tau + \dot{p}_\tau\left(\left[\frac{1}{1+e^{V_\tau}} - p_\tau\right]\sinh(V_\tau) - \frac{1}{2}\right)^{-1}}\left[\dot{p}_\tau V_\tau - P\right] = \dot{p}_\tau \ln\left(\frac{p_\tau}{1-p_\tau}\right) + \frac{[\Delta H + S_\epsilon]}{\tau}.$$

The combined constraint becomes

$$F = \frac{\dot{p}_\tau \ln\left(\frac{p_\tau}{1-p_\tau}\right) + \frac{[\Delta H + S_\epsilon]}{\tau}}{[\dot{p}_\tau V_\tau - P]} - \frac{\ln\left(\frac{p_\tau}{1-p_\tau}\right)}{V_\tau + \dot{p}_\tau\left(\left[\frac{1}{1+e^{V_\tau}} - p_\tau\right]\sinh(V_\tau) - \frac{1}{2}\right)^{-1}} = 0 \tag{A.38}$$

Bibliography

[1] W. Nernst, "Ueber die berechnung chemischer gleichgewichte aus thermischen messungen," *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, vol. 1906, pp. 1–40, 1906.

[2] A. Einstein, "Die plancksche theorie der strahlung und die theorie der spezifischen waerme," *Annalen der Physik*, vol. 327, no. 1, pp. 180–190, 1907.

[3] M. Planck, *Thermodynamik*. Walter de Gruyter and Company, 1954.

[4] B. Cleuren, B. Rutten, and C. Van den Broeck, "Cooling by heating: Refrigeration powered by photons," *Physical review letters*, vol. 108, no. 12, p. 120603, 2012.

[5] R. Fox, *Caloric theory of gases*. Clarendon Press, 1971.

[6] C. G. Knott, *Life and scientific work of Peter Guthrie Tait*, vol. 2. Ripol Classic Publishing House, 1911.

[7] L. Szilard, "über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen," *Zeitschrift für Physik*, vol. 53, pp. 840–856, Nov 1929.

[8] M. Smoluchowski, "Experimentell nachweisbare, der üblichen thermodynamik widersprechende molekularphänomene," *Pisma Mariana Smoluchowskiego*, vol. 2, no. 1, pp. 226–251, 1927.

[9] R. P. Feynman, R. B. Leighton, and M. Sands, "The feynman lectures on physics; vol. i," *American Journal of Physics*, vol. 33, no. 9, pp. 750–752, 1965.

[10] J. Norton, "All shook up: fluctuations, maxwell's demon and the thermodynamics of computation," *Entropy*, vol. 15, no. 10, pp. 4432–4483, 2013.

[11] G. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, "Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales," *Physical Review Letters*, vol. 89, no. 5, p. 050601, 2002.

[12] T. M. NIEUWENHUIZEN and A. E. ALLAHVERDYAN, "Comment on" experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales"," *Fluctuation and Noise Letters*, vol. 5, no. 02, pp. C23–C24, 2005.

[13] D. J. Evans and D. J. Searles, "The fluctuation theorem," *Advances in Physics*, vol. 51, no. 7, pp. 1529–1585, 2002.

[14] J. Kurchan, "Fluctuation theorem for stochastic dynamics," *Journal of Physics A: Mathematical and General*, vol. 31, no. 16, p. 3719, 1998.

[15] G. E. Crooks, "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences," *Physical Review E*, vol. 60, no. 3, p. 2721, 1999.

[16] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco Jr, and C. Bustamante, "Verification of the crooks fluctuation theorem and recovery of rna folding free energies," *Nature*, vol. 437, no. 7056, p. 231, 2005.

[17] C. H. Bennett, "The thermodynamics of computation—a review," *International Journal of Theoretical Physics*, vol. 21, no. 12, pp. 905–940, 1982.

[18] O. Maroney, "Information processing and thermodynamic entropy," in *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.), Metaphysics Research Lab, Stanford University, fall 2009 ed., 2009.

[19] L. Brillouin, "Maxwell's demon cannot operate: Information and entropy. i," *Journal of Applied Physics*, vol. 22, no. 3, pp. 334–337, 1951.

[20] L. Brillouin, *Science and information theory*. Courier Corporation, 2013.

[21] D. Gabor, "Iv light and information," in *Progress in optics*, vol. 1, pp. 109–153, Elsevier, 1961.

[22] J. Rothstein, "Information, measurement, and quantum mechanics," *Science*, vol. 114, no. 2955, pp. 171–175, 1951.

[23] C. H. Bennett, "Demons, engines and the second law," *Scientific American*, vol. 257, no. 5, pp. 108–117, 1987.

[24] C. H. Bennett, "Logical reversibility of computation," *IBM journal of Research and Development*, vol. 17, no. 6, pp. 525–532, 1973.

[25] C. H. Bennett, "Notes on landauer's principle, reversible computation, and maxwell's demon," *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, vol. 34, no. 3, pp. 501–510, 2003.

[26] R. Landauer, "Irreversibility and heat generation in the computing process," *IBM journal of research and development*, vol. 5, no. 3, pp. 183–191, 1961.

[27] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[28] E. T. Jaynes, *Probability theory: The logic of science*. Cambridge university press, 2003.

[29] T. Sagawa, "Thermodynamic and logical reversibilities revisited," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2014, p. P03025, mar 2014.

[30] R. Landauer, "Minimal energy requirements in communication," *Science*, vol. 272, no. 5270, pp. 1914–1918, 1996.

[31] R. Landauer, "The physical nature of information," *Physics letters A*, vol. 217, no. 4-5, pp. 188–193, 1996.

[32] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, "Experimental verification of landauer's principle linking information and thermodynamics," *Nature*, vol. 483, no. 7388, p. 187, 2012.

[33] Y. Jun, M. Gavrilov, and J. Bechhoefer, "High-precision test of landauer's principle in a feedback trap," *Physical review letters*, vol. 113, no. 19, p. 190601, 2014.

[34] A. O. Orlov, C. S. Lent, C. C. Thorpe, G. P. Boechler, and G. L. Snider, "Experimental test of landauer's principle at the sub-kbt level," *Japanese Journal of Applied Physics*, vol. 51, no. 6S, p. 06FE10, 2012.

[35] G. N. Price, S. T. Bannerman, K. Viering, E. Narevicius, and M. G. Raizen, "Single-photon atomic cooling," *Phys. Rev. Lett.*, vol. 100, p. 093004, Mar 2008.

[36] M. G. Raizen, "Comprehensive control of atomic motion," *Science*, vol. 324, no. 5933, pp. 1403–1406, 2009.

[37] J. J. Thorn, E. A. Schoene, T. Li, and D. A. Steck, "Experimental realization of an optical one-way barrier for neutral atoms," *Phys. Rev. Lett.*, vol. 100, p. 240407, Jun 2008.

[38] 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," *Nature physics*, vol. 6, no. 12, p. 988, 2010.

[39] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, "Experimental verification of landauer's principle linking information and thermodynamics," *Nature*, vol. 483, no. 7388, p. 187, 2012.

[40] V. Serreli, C.-F. Lee, E. R. Kay, and D. A. Leigh, "A molecular information ratchet," *Nature*, vol. 445, p. 523, 2007.

[41] J. V. Koski, V. F. Maisi, J. P. Pekola, and D. V. Averin, "Experimental realization of a szilard engine with a single electron," *Proceedings of the National Academy of Sciences*, vol. 111, no. 38, pp. 13786–13789, 2014.

[42] J. V. Koski, A. Kutvonen, I. M. Khaymovich, T. Ala-Nissila, and J. P. Pekola, "On-Chip Maxwell's Demon as an Information-Powered Refrigerator," *Phys. Rev. Lett.*, vol. 115, p. 260602, Dec 2015.

[43] K. Chida, K. Nishiguchi, G. Yamahata, H. Tanaka, and A. Fujiwara, "Thermal-noise suppression in nano-scale si field-effect transistors by feedback control based on single-electron detection," *Applied Physics Letters*, vol. 107, no. 7, 2015.

[44] M. D. Vidrighin, O. Dahlsten, M. Barbieri, M. Kim, V. Vedral, and I. A. Walmsley, "Photonic maxwell's demon," *Physical review letters*, vol. 116, no. 5, p. 050401, 2016.

[45] H. van Houten, C. Beenakker, and B. Van Wees, "Quantum point contacts," in *Semiconductors and Semimetals*, vol. 35, pp. 9–112, Elsevier, 1992.

[46] A. Hosoya, K. Maruyama, and Y. Shikano, "Maxwell's demon and data compression," *Physical Review E*, vol. 84, no. 6, p. 061117, 2011.

[47] M. Born, "Quantenmechanik der stoßvorgänge," *Zeitschrift für Physik*, vol. 38, no. 11-12, pp. 803–827, 1926.

[48] J. Von Neumann, *Mathematische grundlagen der quantenmechanik*, vol. 38. Springer-Verlag, 2013.

[49] K. Jacobs, *Quantum measurement theory and its applications*. Cambridge University Press, 2014.

[50] S. W. Kim, T. Sagawa, S. De Liberato, and M. Ueda, "Quantum szilard engine," *Phys. Rev. Lett.*, vol. 106, p. 070401, Feb 2011.

[51] T. D. Kieu, "The second law, maxwell's demon, and work derivable from quantum heat engines," *Phys. Rev. Lett.*, vol. 93, p. 140403, Sep 2004.

[52] V. B. Sørdal and J. Bergli, "Quantum particle in a split box: Excitations to the ground state," *Physical Review A*, vol. 99, no. 2, p. 022121, 2019.

[53] W. H. Zurek, *Maxwell's demon, Szilard's engine and quantum measurements*. Springer, 1986.

[54] W. H. Zurek, *Information transfer in quantum measurements: Irreversibility and amplification*. Springer, 1983.

[55] L. L. Yan, T. P. Xiong, K. Rehan, F. Zhou, D. F. Liang, L. Chen, J. Q. Zhang, W. L. Yang, Z. H. Ma, and M. Feng, "Single-atom demonstration of the quantum landauer principle," *Phys. Rev. Lett.*, vol. 120, p. 210601, May 2018.

[56] A. Géron, *Hands-on Machine Learning with Scikit-Learn and Tensorflow*. O'Reilly Media, Inc.

[57] J. F. Trevor Hastie, Robert Tibshirani, *The Elements of Statistical Learning: Data Mining, Inferene, and Prediction*. Springer Science + Business Media.

[58] R. BELLMAN, "A markovian decision process," *Journal of Mathematics and Mechanics*, vol. 6, no. 5, pp. 679–684, 1957.

[59] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, Feb. 2015.

[60] S. S. Patel and I. Donmez, "Mechanisms of helicases," *Journal of Biological Chemistry*, vol. 281, no. 27, pp. 18265–18268, 2006.

[61] J. M. Caruthers and D. B. McKay, "Helicase structure and mechanism," *Current opinion in structural biology*, vol. 12, no. 1, pp. 123–133, 2002.

[62] A. Ghosh and M. Bansal, "A glossary of dna structures from a to z," *Acta Crystallographica Section D: Biological Crystallography*, vol. 59, no. 4, pp. 620–626, 2003.

[63] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, "Stretching dna with optical tweezers," *Biophysical journal*, vol. 72, no. 3, pp. 1335–1346, 1997.

[64] F. Ritort, "Nonequilibrium fluctuations in small systems: From physics to biology," 2008.

[65] D. V. Averin and J. P. Pekola, "Reversing the landauer's erasure: Single-electron maxwell's demon operating at the limit of thermodynamic efficiency," *physica status solidi (b)*, vol. 254, no. 3, p. 1600677, 2017.

Papers

1. Cooling by heating: Restoration of the third law of thermodynamics 2016
V.B. Sørdal, J. Bergli, Y.M. Galperin *Physical Review E 93 (3), 032102*

2. Influence of measurement error on Maxwell's demon 2017
V.B. Sørdal, J. Bergli, Y.M. Galperin *Physical Review E 95 (6), 062129*

3. Quantum particle in a split box: Excitations to the ground state 2019
V.B. Sørdal, J. Bergli *Physical Review A 99 (2), 022121*

4. Deep reinforcement learning for robust quantum optimization 2019
V.B. Sørdal, J. Bergli *Submitted to Physical Review A*

I

Cooling by heating: Restoration of the third law of thermodynamics

V. B. Sørdal,[1,*] J. Bergli,[1] and Y. M. Galperin[1,2]
[1]*Department of Physics, University of Oslo, P.O. Box 1048 Blinderm, 0316 Oslo, Norway*
[2]*Ioffe Institute, 26 Politekhnicheskaya, St. Petersburg 194021, Russian Federation*
(Received 20 November 2015; published 2 March 2016)

We have made a simple and natural modification of a recent quantum refrigerator model presented by Cleuren *et al.* [Phys. Rev. Lett. **108**, 120603 (2012)]. The original model consist of two metal leads acting as heat baths and a set of quantum dots that allow for electron transport between the baths. It was shown to violate the dynamic third law of thermodynamics (the unattainability principle, which states that cooling to absolute zero in finite time is impossible). By taking into consideration the finite energy level spacing Δ, in metals we restore the third law while keeping all of the original model's thermodynamic properties intact down to the limit of $k_BT \sim \Delta$, where the cooling rate is quenched. The spacing Δ depends on the confinement of the electrons in the lead and therefore, according to our result larger samples (with smaller level spacing), could be cooled efficiently to lower absolute temperatures than smaller ones. However, a large lead makes the assumption of instant equilibration of electrons implausible; in reality one would only cool a small part of the sample and we would have a nonequilibrium situation. This property is expected to be model independent and raises the question whether we can find an optimal size for the lead that is to be cooled.

DOI: 10.1103/PhysRevE.93.032102

I. INTRODUCTION

Quantum refrigerators are solid-state devices with huge potential benefits in technology. With no moving parts and of microscopic size, they could easily be integrated into existing technology, such as cellphones and computers, to enhance their performance by utilizing the waste heat energy they produce. As always, the technological frontier is supported by a backbone of theoretical framework, which in recent years has seen many advancements (see, e.g., [1–6]). In addition to the technological possibilities they present, quantum refrigerators are excellent tools for providing insight into the unique features of open quantum systems. For a review of stochastic thermodynamics and the formalism used to treat quantum refrigerators see, e.g., [7,8].

The quantum absorption refrigerator is a version of these general machines, based on producing a steady-state heat flow from a cold to a hot reservoir, driven by absorption from an external heat reservoir. A key tool to understand the operation of these refrigerators, when approaching the limiting temperature of absolute zero, is the laws of thermodynamics. In this article we study one such device that appears to violate the dynamic version of the third law of thermodynamics (the unattainability principle), which states that one cannot cool a system to absolute zero in a finite amount of time. A recent publication by Cleuren *et al.* [9] presented a novel model based on two electronic baths coupled together via a system of quantum dots and driven by an external photon source. The article generated some controversy due to its apparent violation of the unattainability principle, and several authors [10–13] proposed explanations for this violation. However, we find that the discussion was without conclusion, and we will discuss this later in the article.

We will begin by giving a brief presentation of the quantum refrigerator model, as introduced by Cleuren *et al.* [9], and its thermodynamic properties. Then we will summarize and comment on the discussion that followed. Finally, we will present a simple modification, based only on the fact that the energy levels of metals are discrete when treated quantum mechanically, which becomes important at temperatures $T \lesssim \Delta$, where Δ is the level spacing. (We measure temperature in energy units, setting the Boltzmann constant $k_B = 1$.) Our modification upholds the third law while it simultaneously reproduces the results from the original model down to the limit of $T \sim \Delta$. In essence, we want to make the point that the validity of the unattainability principle is only guaranteed when applied to a quantum description of a system and that the most important quantum effect to consider in relation to this law is the discretization of energy states.

A. Model

The quantum refrigerator model proposed in [9] is shown schematically in real space in Fig. 1 and in energy space in Fig. 2. Here we briefly explain its operating protocol. It consists of two metal leads and four quantum dots; the large and hot lead with temperature T_L is coupled to the small cold lead with temperature T_R, via the set of quantum dots. We assume that each quantum dot is highly confined and is thus associated with a single energy level, since the other levels are far outside the energy range of the system. These four levels are marked in Fig. 2. The quantum dots form two channels, as illustrated in Fig. 1, where the energy levels ϵ_2 (ϵ_1) and $\epsilon_2 + \epsilon_g$ ($\epsilon_1 - \epsilon_g$) are coupled together in channel 2 (channel 1). The two channels are spatially separated, therefore we can safely ignore any Coulomb interaction between the electrons in channels 1 and 2. The basic idea is to move cold electrons (i.e., with energy less than μ) from the hot lead into the cold lead via channel 1 while simultaneously moving hot electrons (energy greater than the chemical potential μ) from the cold lead to the hot lead via channel 2. This transport of electrons will thus cool the right lead by injecting cold and extracting hot electrons. Naturally the transport will also heat up the left lead, but since we assume

*v.b.sordal@fys.uio.no

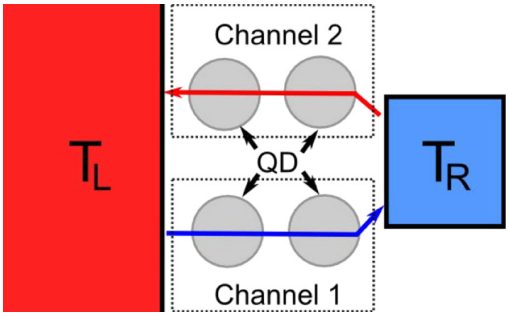

FIG. 1. Schematic of the model shown in real space. A small piece of metal with temperature T_R is coupled to a larger piece with temperature $T_L > T_R$. Four quantum dots form two channels for electron transport between the metals. The arrows indicate the desired direction of the net particle current to achieve cooling of the right metal lead. The distance between the two channels is too large for any Coulomb interaction to take place between them.

that it is a large piece of metal with a high heat capacity, the heat absorbed will not result in a measurable change in T_L. We can obtain the desired particle flow direction by coupling the quantum dot system to a bosonic bath that induces transitions between the quantum dots of each pair, i.e., between ϵ_1 and $\epsilon_1 - \epsilon_g$ in channel 1 and between ϵ_2 and $\epsilon_2 + \epsilon_g$ in channel 2. The bosonic bath can be photons from an external source and/or phonons from the device. In this discussion we will consider it to be a photon bath with temperature T_S. In Ref. [9] the photon bath is taken to be the sun with a temperature $T_S \simeq 6000$ K and we will follow this in the sense that we will assume that it is the largest energy scale in the system. In any case the transition rates between the quantum dots are proportional to the probability of finding a boson with energy equal to the energy difference between the two quantum dot levels, which is given by the Planck distribution $n(E)$. The

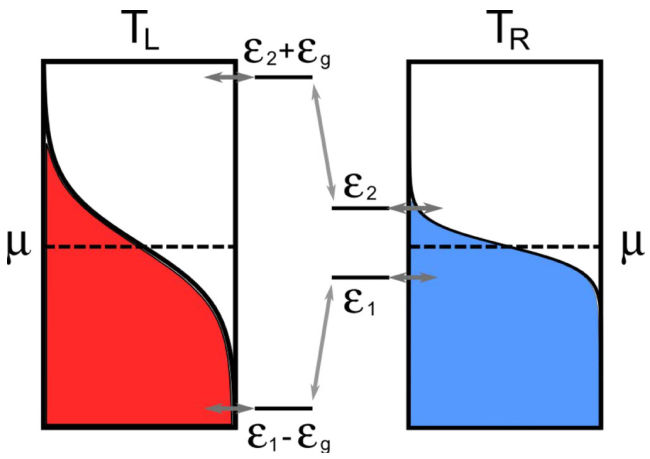

FIG. 2. A hot metal lead T_L is coupled to a cold one T_R via two spatially separated pairs of quantum dots, which form two channels for electron transport between the leads. We consider the case where $\mu_L = \mu_R = \mu$ and the energy levels of the quantum dots are symmetric about the chemical potential ($\epsilon_2 - \mu = \mu - \epsilon_1 \rightarrow \epsilon_1 = -\epsilon_2$). The schematic is adapted from [9].

rates are thus given by

$$k_{\uparrow}^{\epsilon_g} = \frac{\Gamma_s}{e^{\epsilon_g/T_S} - 1}, \quad k_{\downarrow}^{\epsilon_g} = \frac{\Gamma_s}{1 - e^{-\epsilon_g/T_S}}. \tag{1}$$

Here $k^{\uparrow}$ and $k^{\downarrow}$ are the rates for upward and downward transitions in energy, respectively. The difference between them is that $k^{\downarrow}$ contains an additional term for spontaneous emission.

The transition rate for electron transfer from the metal to an empty quantum dot level is proportional to the probability of finding an electron in the same energy level in the metal, which is given by the Fermi-Dirac distribution $f(E)$. For the inverse transition to take place there has to be an available energy level in the metal, which has a probability proportional to $1 - f(E)$. Thus the transition rates between the quantum dot and metal are

$$k_{l\to d}^{E} = \frac{\Gamma}{e^{(E-\mu)/T} + 1}, \quad k_{d\to l}^{E} = \frac{\Gamma}{e^{(\mu-E)/T} + 1}. \tag{2}$$

For transitions involving the right lead the temperature $T = T_R$, while for the left lead $T = T_L$. Notice that in general $\Gamma \neq \Gamma_s$. These are the constants that set the time scale of the transitions and depends on the specific details of the device.

As in Ref. [9], we will considering the strongly coupled case where the energies of the quantum dots are symmetric about the chemical potential ($\epsilon_2 - \mu = \mu - \epsilon_1$). We can therefore choose to measure all energies relative to $\mu = 0$ and combine the two parameters $\epsilon_2 = -\epsilon_1 = \epsilon$.

We can now introduce three distinct occupation probabilities per channel. Since the two quantum dots in the same channel are close to each other in space we assume that the Coulomb repulsion between electrons prevents simultaneous occupation of the right and left quantum dots. For channel 1 we then have the probabilities $P_L^{(1)}$, $P_R^{(1)}$, and $P_0^{(1)}$, which represent the probability of finding an electron in the left quantum dot with energy $-(\epsilon + \epsilon_g)$, in the left quantum dot with energy $-\epsilon$, and in neither quantum dot, respectively. A master equation describing the time evolution of the occupation probabilities in channel 1 can thus be formulated

$$\dot{\mathbf{P}}^{(1)} = \hat{M}^{(1)}\mathbf{P}^{(1)}, \quad \mathbf{P}^{(1)} \equiv \begin{bmatrix} P_0^{(1)} \\ P_L^{(1)} \\ P_R^{(1)} \end{bmatrix}, \tag{3}$$

where the transition matrix $M^{(1)}$ is given by

$$M^{(1)} = \begin{bmatrix} -k_{l\to d}^{-(\epsilon+\epsilon_g)} - k_{l\to d}^{-\epsilon} & k_{d\to l}^{-(\epsilon+\epsilon_g)} & k_{d\to l}^{-\epsilon} \\ k_{l\to d}^{-(\epsilon+\epsilon_g)} & -k_{d\to l}^{-(\epsilon+\epsilon_g)} - k_{\uparrow}^{\epsilon_g} & k_{\downarrow}^{\epsilon_g} \\ k_{l\to d}^{-\epsilon} & k_{\uparrow}^{\epsilon_g} & -k_{d\to l}^{-\epsilon} - k_{\downarrow}^{\epsilon_g} \end{bmatrix}.$$

We are interested in the steady state of the system, where the probabilities do not change as a function of time. To find this state we set $\dot{\mathbf{P}}^{(1)} = \mathbf{0}$ and solve Eq. (3). By doing this we obtain the steady-state probability vector $\mathbf{P}^{(1)}(\epsilon, \epsilon_g, T_R, T_L)$, where we consider Γ, Γ_s, and T_S as constants. A similar procedure gives us the steady-state probability vector for channel 2 as well.

The particle current between the right dot in the lower level and the cold lead can be written as

$$J^{(1)} = P_R^{(1)} k_{d\to l}^{-\epsilon} - P_0^{(1)} k_{l\to d}^{-\epsilon} \tag{4}$$

and the current through the upper level is

$$J^{(2)} = P_R^{(2)} k_{d\to l}^{\epsilon} - P_0^{(2)} k_{l\to d}^{\epsilon}. \tag{5}$$

The cooling power, i.e., the heat transported *out of* the right lead per unit time, can now be defined as

$$\dot{Q}_R = (-\epsilon - \mu)(-J^{(1)}) + (\epsilon - \mu)(-J^{(2)}). \tag{6}$$

Since the energy levels are symmetric about μ, we can set $\mu = 0$ and we obtain the cooling power for the refrigerator model

$$\dot{Q}_R = \epsilon(J^{(1)} - J^{(2)}). \tag{7}$$

Optimized cooling is attained by varying $\epsilon(T_R)$ and $\epsilon_g(T_R)$ as a function of T_R (when T_S and T_L are kept constant). It can be shown (see Ref. [9] for details) that the cooling power in the limit of low T_R is given by

$$\lim_{T_R\to 0} \dot{Q}_R \propto T_R. \tag{8}$$

When working at an energy scale where $\epsilon_g \ll T_S$ we have $k^{\uparrow} \simeq k^{\downarrow}$. In this situation we can get a better understanding of the system and when cooling will occur by considering the transitions in channel 2. There the energy levels are situated above μ and we have

$$0 < f(E) < 1/2, \quad 1/2 < 1 - f(E) < 1.$$

Therefore, the rate from lead to dot will always be less than the rate from dot to lead $k_{l\to d}^{E} < k_{d\to l}^{E}$ for a given energy E. The requirement for cooling to take place in this situation is that $f(\epsilon + \epsilon_g, T_L) < f(\epsilon, T_R)$, i.e., we require $(\epsilon + \epsilon_g)/T_L > \epsilon/T_R$. We then have

$$\left.\begin{array}{l} k_{d\to l}^{\epsilon+\epsilon_g} > k_{d\to l}^{\epsilon} \\ k_{l\to d}^{\epsilon+\epsilon_g} < k_{l\to d}^{\epsilon} \\ k_{l\to d}^{E} < k_{d\to l}^{E} \end{array}\right\} \Rightarrow k_{d\to l}^{\epsilon+\epsilon_g} > k_{d\to l}^{\epsilon} > k_{l\to d}^{\epsilon} > k_{l\to d}^{\epsilon+\epsilon_g}. \tag{9}$$

When $k^{\uparrow} \simeq k^{\downarrow}$ we know that the occupation probability $P_L^{(2)} \simeq P_R^{(2)} = P$ and thus $P_0^{(2)} = 1 - 2P$. Using the inequalities shown in Eq. (9), we now consider two different states of the system. First assume that there is an electron in the quantum dot system; it can exit into either the left lead or the right lead, where the currents are $k_{d\to l}^{\epsilon+\epsilon_g} P_L^{(2)}$ and $k_{d\to l}^{\epsilon} P_R^{(2)}$, respectively. The difference is

$$P\left(k_{d\to l}^{\epsilon+\epsilon_g} - k_{d\to l}^{\epsilon}\right) > 0,$$

which tells us that it is more likely for the electron to exit into the left lead. Next we assume that the quantum dot system is unoccupied; an electron can enter from the left lead or the right lead, with currents $k_{l\to d}^{\epsilon+\epsilon_g} P_0^{(2)}$ and $k_{l\to d}^{\epsilon} P_0^{(2)}$, respectively. The difference is now

$$(1 - 2P)\left(k_{l\to d}^{\epsilon+\epsilon_g} - k_{l\to d}^{\epsilon}\right) < 0,$$

indicating that it is more likely that an electron enters from the right lead. Above the chemical potential, electrons entering from the right lead and exiting into the left lead correspond to a net cooling of the right lead, which is our desired effect. A similar analysis can be done for channel 1, where the corresponding result of net transport from the left to the right lead is obtained.

In summary, one obtains optimal cooling of the right lead by varying the energy levels $\epsilon_2 = -\epsilon_1 = \epsilon$ as a function of T_R and their optimal position is determined by a balance between the transport rate (higher closer to μ) and heat removed per transition (higher far from μ), and the additional requirement that $f(\epsilon + \epsilon_g, T_L) < f(\epsilon, T_R)$.

B. Unattainability principle

The unattainability principle states that one cannot cool a system to absolute zero in a finite amount of time [14]. A system with heat capacity $C_V = dQ/dT$ and cooling power $\dot{Q} = dQ/dt$ has a cooling rate given by

$$\frac{dT}{dt} = \frac{\dot{Q}}{C_V}. \tag{10}$$

If we assume that C_V and $\dot{Q}$ scale with temperature to the power of κ and λ, respectively, we have

$$\frac{dT}{dt} \propto T^{\lambda-\kappa}. \tag{11}$$

For $\alpha \equiv \lambda - \kappa < 1$ the unattainability principle is violated [10] and cooling to absolute zero is possible in a finite time. By inspecting Eq. (8) we find that $\lambda = 1$. The heat capacity of the metal lead as $T_R \to 0$ is dominated by the electronic heat capacity, which is proportional to the temperature $C_V \propto T_R$ (see, e.g., Ref. [15]), and therefore $\kappa = 1$. The end result is that $\dot{T}_R \propto T^0$, in violation of the unattainability principle.

C. Comments

Levy *et al.* [10] were the first to point out that because the refrigerator presented in [9] has a cooling power of $\dot{Q} \propto T_R$ and a heat capacity of $C_V \propto T_R$ in the limit of $T_R \to 0$ K, its cooling rate is given by

$$\frac{dT(t)}{dt} = \frac{\dot{Q}}{C_V} \propto T_R^0 = \text{const.} \tag{12}$$

That enables cooling to absolute zero in a finite amount of time. In the original model proposed by Cleuren *et al.* the quantum dot system consisted of only two quantum dots, with the levels ϵ_1 $(\epsilon_1 - \epsilon_g)$ and ϵ_2 $(\epsilon_2 + \epsilon_g)$ being two adjacent levels within the right (left) quantum dot. Levy *et al.* suggest that the violation of the third law may be due to the neglect of internal transitions within a single dot. This suggestion was refuted by Cleuren *et al.* [11], who stated that the model could also be constructed using two pairs of spatially separated quantum dots, as we have done here. Their own explanation for the violation was that the quantum master equation they utilized does not take into account coherent effects and the broadening of the linewidth of the quantum dot energy levels was ignored. Both of these effects becomes important in the low-temperature limit.

Allahverdyan *et al.* [12] suggested that the violation occurs since the weak-coupling master equation used by Cleuren *et al.* is limited at low temperatures. They state that one can justify taking the limit $T_R \to 0$ for such an equation only while

simultaneously reducing the coupling between the quantum dot system and heat reservoirs $\gamma \to 0$. Concrete analysis of the low-temperature behavior of the cooling power is not given.

Finally, Entin-Wohlman and Imry [13] considered a simplified version of the original model, where only a single channel contributes to the electron transfer. They assumed that boson-assisted hopping is the dominant form of electronic transport [2] (an assumption we will also make later in the article). If we remove channel 1 from our model and only consider channel 2, we obtain the same system as considered in [2]. Using Fermi's golden rule, they found that the heat current is exponentially small for $\epsilon_2 - \mu \gg T_R$. They went on to state that the violation of the third law comes from allowing the levels ϵ_1 and ϵ_2 to approach the chemical potential linearly as a function of temperature and claimed that this is unnecessary and complicates the setup. In our opinion, the linear temperature dependence of the energy levels ϵ_1 and ϵ_2 in the quantum dots coupled to the cold lead is an essential feature; it arises from the optimization of the cooling power suggested in Ref. [9], but not implemented in Ref. [13].

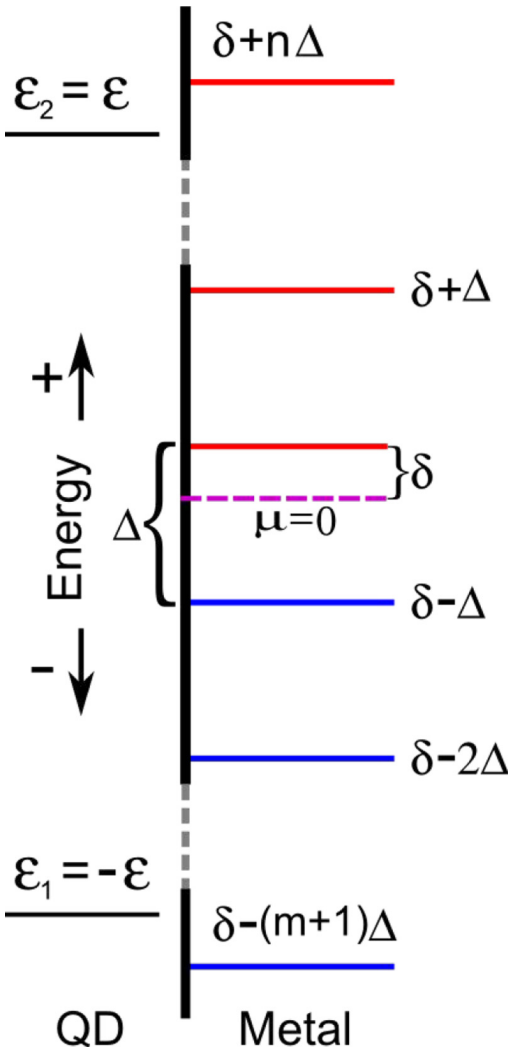

FIG. 3. The continuous states of the metal are replaced by a discrete spectrum with a constant energy spacing Δ. The asymmetry between states above and below μ is modeled by the parameter δ. For $\delta = \Delta/2$ the chemical potential lies exactly in the middle of two energy levels. The jth (ith) level below (above) μ is given by $\epsilon_j = \delta - j\Delta$ [$\epsilon_i = \delta + (i-1)\Delta$].

II. DISCRETIZATION OF THE MODEL

One of the assumptions of the model proposed is that there is a continuous spectrum of energy states in the metal leads. Thus the electrons are transferred elastically between the quantum dots and the metals. We will now introduce a simple discretized modification of the original model and show that the unattainability principle will then be restored. In our model, we assume an even spacing between the energy levels. We also introduce the parameter δ to quantify the asymmetry about the chemical potential μ (see Fig. 3). If $\delta = \Delta/2$ the energy levels are symmetrically distributed about μ. As long as the quantum dot and metal energy levels do not exactly overlap, the transitions are now inelastic and require absorption or emission of phonons.

A. Cooling power

We can set up a master equation for the dynamics in channel 1, as in Eq. (3), but now for the discrete system. The rate matrix is almost identical, but since we allow for phonon-assisted transitions, the rates between the quantum dots and the discrete levels of the right lead are given by a sum of all possible emission and absorption transitions. We will use ϵ_n (ϵ_m) to denote the nth (mth) level in the metal lead, above (below) the quantum dot level ϵ_1. We also introduce $\omega_n = \epsilon_n - \epsilon_1$ and $\omega_m = \epsilon_1 - \epsilon_m$ to represent the phonon frequencies associated with transitions between these levels. For transitions from the lead to the dot, ϵ_n and ϵ_m are the energies associated with emission and absorption processes, respectively, while for dot-to-lead transitions the association is opposite. The matrix elements change from $k^{\epsilon_1}_{d\to l} \to k^{d,\epsilon_1}_{d\to l}$ and $k^{\epsilon_1}_{l\to d} \to k^{d,\epsilon_1}_{l\to d}$, where we use the superscript d to indicate that it is the transition rate for the discrete model. These rates are then sums of all possible emission and absorption processes and can be written as

$$\begin{aligned} k^{d,\epsilon_1}_{d\to l} &= \overbrace{\sum_m k^{\epsilon_m}_{d\to l}}^{\text{emission}} + \overbrace{\sum_n k^{\epsilon_n}_{d\to l}}^{\text{absorption}}, \\ k^{d,\epsilon_1}_{l\to d} &= \underbrace{\sum_m k^{\epsilon_m}_{l\to d}}_{\text{absorption}} + \underbrace{\sum_n k^{\epsilon_n}_{l\to d}}_{\text{emission}}, \end{aligned} \tag{13}$$

where the emission and absorption rates are given by

$$\begin{aligned} k^{\epsilon_n}_{d\to l} &= \Gamma[1 - f(\epsilon_n)]n(\omega_n)\omega_n^2, \\ k^{\epsilon_m}_{d\to l} &= \Gamma[1 - f(\epsilon_m)][n(\omega_m) + 1]\omega_m^2, \\ k^{\epsilon_m}_{l\to d} &= \Gamma f(\epsilon_m)n(\omega_m)\omega_m^2, \\ k^{\epsilon_n}_{l\to d} &= \Gamma f(\epsilon_n)[n(\omega_n) + 1]\omega_n^2. \end{aligned} \tag{14}$$

Here $n(\omega) = (e^{\omega/T_R} - 1)^{-1}$ is the Planck distribution, which tells us the probability of finding a phonon with energy ω, and $f(\epsilon) = (e^{\epsilon/T_R} + 1)^{-1}$ is the Fermi-Dirac distribution, which tells us the probability of finding an occupied state at ϵ. We assume a three-dimensional phonon density of states, thus the

rates have to be multiplied by a ω^2 term. We have absorbed all other constants from the density of states into the Γ introduced earlier.

The transitions between the left quantum dot and the hot left lead, i.e., the rates involving $-(\epsilon+\epsilon_g)$, remain unchanged since we still consider this to be a large metal piece with a quasicontinuous energy spectrum. Again, we solve the master equation in the steady state and obtain the occupation probability vector $\mathbf{P}^{(1)}$, but now for the discrete model. With this we can find the particle currents in the channel 1 for the discrete model,

$$J_d^{(1)} = P_R^{(1)} k_{d\to l}^{d,\epsilon_1} - P_0^{(1)} k_{l\to d}^{d,\epsilon_1}. \tag{15}$$

Thus we can write the part of the cooling power associated with channel 1 as

$$\dot{Q}_R^{(1)} = P_R^{(1)}\left(\sum_m k_{d\to l}^{\epsilon_m}\epsilon_m + \sum_n k_{d\to l}^{\epsilon_n}\epsilon_n\right) - P_0^{(1)}\left(\sum_m k_{l\to d}^{\epsilon_m}\epsilon_m + \sum_n k_{l\to d}^{\epsilon_n}\epsilon_n\right). \tag{16}$$

An analysis similar to that shown here can be applied to channel 2 and provide its corresponding cooling power $\dot{Q}_R^{(2)}$. Thus the total cooling power is written as

$$\dot{Q}_R = \dot{Q}_R^{(1)} + \dot{Q}_R^{(2)}. \tag{17}$$

It should be noted that in the limit of $T_R \to 0$ only the two levels δ and $\delta-\Delta$ will contribute to the total cooling power since all levels above δ will be unoccupied and all levels below $\delta-\Delta$ will be occupied. We can now numerically optimize Eq. (17), with respect to the two parameters ϵ and ϵ_g, while keeping T_L and T_S constant. Note that ϵ_m and ϵ_n are determined from $\epsilon = -\epsilon_1 = \epsilon_2$ and are not free parameters. When $\epsilon_g \gg \epsilon$, the optimal energy of the quantum dot levels $\pm(\epsilon_g+\epsilon)$ is independent of ϵ and therefore also independent of T_R (the only influence of T_R on those levels come via the coupling to the levels $\pm\epsilon$). This in turn makes the optimal cooling power $\dot{Q}_R$ approximately independent of ϵ_g. Hence the only free parameter for optimization is $\epsilon(T_R)$.

The plot of the optimized cooling power as a function of T_R is shown in Fig. 4. As in [9], we have to use numerics to analyze the behavior of the optimized $\dot{Q}_R$. For simplicity we set $\delta = \Delta/2$ and by fitting the numerical results from Eq. (17) to the Arrhenius equation ($\ln\dot{Q} = \ln A - B/T$), we find that the optimized cooling power as $T_R \to 0$ K is given by

$$\dot{Q}_R \propto e^{-\Delta/2T_R}, \quad T_R \to 0. \tag{18}$$

B. Heat capacity

The heat capacity of a Fermi gas with temperature-independent chemical potential μ can be expressed as

$$C_V = \int_0^\infty d\epsilon(\epsilon-\mu)D(\epsilon)\frac{\partial f(\epsilon)}{\partial T}. \tag{19}$$

Here $D(\epsilon)$ is the density of states (which is a constant in our case) and $f(\epsilon)$ is the Fermi-Dirac distribution. When going from the continuous to the discrete description we have to exchange the integral with a sum and the continuous variable

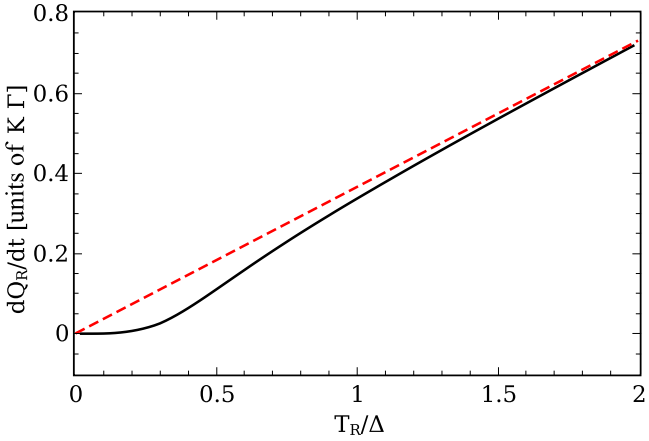

FIG. 4. Graph of the optimized cooling power $\dot{Q}_R$ as a function of the dimensionless variable T_R/Δ. The dashed line is the result from the continuous model, while the solid line is the result from the discrete model. For temperatures $T_R \gtrsim \Delta$ the discrete model reproduces the linear cooling power of the continuous model. However, for temperatures $T_R \lesssim \Delta$ the cooling power changes to an exponential form. The parameters used are $\Gamma = \Gamma_s = 1$, $T_L = 20$ K, $T_S = 6000$ K, $\epsilon_g = 100$ K, $\Delta = 1$ K, and $\delta = \Delta/2$.

ϵ with the discretized states $n\Delta$,

$$C_V = \sum_{n=0}^{\infty}\left(\frac{n\Delta-\mu}{T_R}\right)^2 \frac{e^{(n\Delta-\mu)/T_R}}{(e^{(n\Delta-\mu)/T_R}+1)^2}. \tag{20}$$

This sum can easily be determined numerically, but to gain additional insight we can consider the heat capacity for a two-level system. As $T_R \to 0$ the levels δ and $\delta-\Delta$ will be the only relevant levels. We can write the grand canonical partition function for the two-level system as

$$\Xi = 1 + e^{-\beta\delta} + e^{-\beta(\delta-\Delta)} + e^{-\beta(2\delta-\Delta)} \tag{21}$$

and we can find the energy via

$$U = \frac{1}{\Xi}\sum_i H_i e^{-\beta H_i}, \tag{22}$$

where H_i is the energy of the state i. From this the heat capacity can be obtained from $C_V = dQ/dT = dU/dT$ and we find

$$\begin{aligned} C_V &= -\frac{\Delta^2 A + \delta^2 B + \Delta\delta C}{T_R^2(1+e^{-\beta\delta}+e^{-\beta(\delta-\Delta)}+e^{-\beta(2\delta-\Delta)})^2}, \\ A &= e^{\beta(\Delta-3\delta)} + e^{\beta(\Delta-\delta)} + 2e^{\beta(\Delta-2\delta)}, \\ B &= e^{\beta(\Delta-3\delta)} + e^{\beta(2\Delta-3\delta)} + e^{\beta(\Delta-\delta)} + 4e^{\beta(\Delta-2\delta)} + e^{-\beta\delta}, \\ C &= 2e^{\beta(\Delta-3\delta)} + 2e^{\beta(\Delta-\delta)} + 4e^{\beta(\Delta-2\delta)}. \end{aligned} \tag{23}$$

This expression is greatly simplified at $\delta = \Delta/2$, i.e., a symmetric distribution of energy levels above and below μ. In this case we obtain

$$C_V = 2\left(\frac{\Delta}{2T_R}\right)^2 \frac{e^{\Delta/2T_R}}{(e^{\Delta/2T_R}+1)^2} \tag{24}$$

and with this result, we find that in the limit of $T_R \to 0$ the heat capacity is

$$C_V = 2\left(\frac{\Delta}{2T_R}\right)^2 e^{-\Delta/2T_R}, \quad T_R \to 0. \tag{25}$$

Although this is only true for $\delta = \Delta/2$, we see from the general equation for the heat capacity given in Eq. (23) that the factor of T_R^{-2} is present for all terms and we have found numerically that the dominating exponential terms in the optimized cooling power (17) and the heat capacity (25) always cancel each other as $T_R \to 0$.

III. RESULTS

We can now find the cooling rate dT_R/dt for the discrete system. In Fig. 4 we have plotted the cooling power $\dot{Q}_R$ as a function of the dimensionless variable T_R/Δ. The solid line is the result of our numerical calculations, while the dashed line is the result from the original model [9]. We see that for $T_R \gtrsim \Delta$ the discrete model reproduces the results from the original model, while when $T_R \lesssim \Delta$ the result changes to an exponential form.

The heat capacity C_V is shown as a function of the same dimensionless variable T_R/Δ in the inset in Fig. 4. Again the results from the original model are reproduced for $T_R > \Delta$, but when $T_R < \Delta/2$ a Schottky-like feature appears, indicating that only the two levels closest to $\mu = 0$ are participating in the dynamics.

As we discussed earlier, if we can write the cooling rate in a form like in Eq. (11), we require that $\alpha = \lambda - \kappa \geqslant 1$. In the original model with a continuous energy spectrum in the right metal lead, it was found that $\alpha = 0$. By numerically calculating the expressions given in Eqs. (17) and (20), we find that the cooling rate is given by

$$\frac{dT_R}{dt} \propto \frac{\dot{Q}_R}{C_V} \propto T_R^2, \quad T_R \to 0. \tag{26}$$

We obtain $\alpha = 2$, which implies that cooling to absolute zero is impossible in a finite amount of time, and the discrete model is thus consistent with the unattainability principle. The result is shown in Fig. 5, where we have plotted dT_R/dt as a function of T_R/Δ. Also here the result from the discrete model (solid line) reproduces the result from the original model (dashed line) for $T_R \gtrsim \Delta$, but once $T_R \lesssim \Delta$ it differs.

Although the results from Eqs. (18) and (25) are only valid for $\delta = \Delta/2$, we find numerically that the dominant exponential term in $\dot{Q}_R^{\text{tot}}$ always cancels with the one in C_V. The function $\frac{\dot{T}_R}{T_R^2}$ always converges to a constant value as $T_R \to 0$, thus we conclude that the cooling rate $\dot{T}_R \propto T_R^2$ is valid independent of the choice of δ.

IV. DISCUSSION AND CONCLUSION

We have shown that our natural modification of the model proposed by Cleuren *et at.* does not violate the dynamic version of the third law and allows for the same cooling performance at temperatures $T_R > \Delta$ as the original. This is a positive result, which tells us that the original model can be used to cool very efficiently down to the extreme limit of $T_R \sim \Delta$, where the

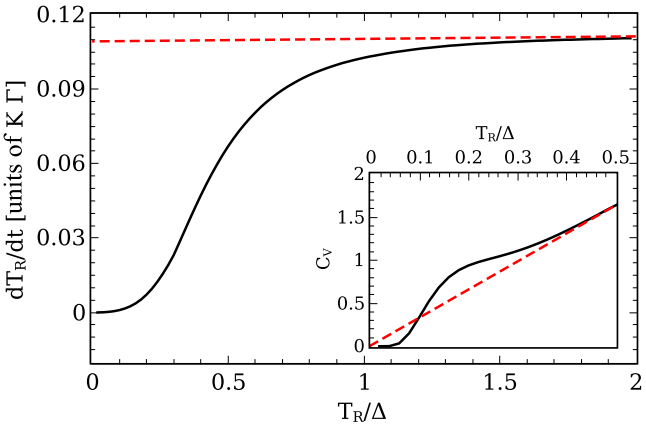

FIG. 5. Plot of the optimized cooling rate dT_R/dt as a function of T_R/Δ, Again we see that the discrete model (solid line) reproduces the third law violating constant rate of temperature change of the continuous model (dashed line) for $T_R \gtrsim \Delta$. The inset shows C_V as a function of the same variable. When $T_R \lesssim \Delta/2$ the heat capacity obtains a feature similar to the Schottky anomaly, indicating that the main contribution to the heat capacity comes from the two levels (δ and $\delta - \Delta$) closest to $\mu = 0$. As a result, for $T_R \lesssim \Delta$ the exponential term in $\dot{Q}_R$ cancels the one in C_V and we are left with the T_R^2 term from the heat capacity. The parameters used are $\Gamma = \Gamma_s = 1$, $T_L = 20$ K, $T_S = 6000$ K, $\epsilon_g = 100$ K, $\Delta = 1$ K, and $\delta = \Delta/2$.

cooling power is quenched. Though we assumed a constant level spacing Δ, the low-temperature behavior of the cooling rate is insensitive to this assumption since at $T_R \to 0$ only the two levels closest to the chemical potential are important.

The laws of thermodynamics are so general that they should apply to both classical and quantum systems. The third law in particular is a theory about the properties of a system as its temperature approaches absolute zero and at low temperatures quantum effects become important. Quantum theory predicts that confined systems have discretized energy levels and when the temperature T becomes comparable to the spacing between energy levels Δ, this discreteness needs to be taken into account. In [9] they use a continuous energy spectrum of the metal lead, disregarding the quantum discreteness. In the comments on the violation of the third law [10–13], they employ a heat capacity derived from quantum theory and it is this mixing of classical and quantum descriptions that leads to the breaking of the unattainability principle. If instead we use a pure classical expression for the heat capacity, which would be a constant as given by the equipartition principle, the unattainability principle would be satisfied [16].

One of the assumptions we have made is that the left hot lead functions as a large heat bath and has no effect on the cooling rate. A recent article [17] has shown that in a cooling process the density of states of the left heat bath affects the cooling rate of quantum refrigerators. A refined model where we take into account the properties of the left lead would give us additional insight into the nature of quantum refrigerators.

We have shown that the cooling power and cooling rate is quenched when $T \sim \Delta$. The energy-level spacing in metals is determined by the strength of confinement of the electrons. By increasing the volume of the lead that is to be cooled, the

spacing Δ will decrease. Therefore, the temperature where the cooling power is quenched approaches absolute zero as the volume goes to infinity. However, by increasing the volume of the lead, the assumption of instantaneous equilibration of the electrons according to the Fermi-Dirac distribution becomes implausible; in reality, only a small area of the sample would be cooled and the lead would be in a nonequilibrium state. Larger leads also have a higher heat capacity and one must remove more heat per degree of temperature change than for smaller samples, which decreases the cooling efficiency. Finding the limit of efficient cooling for real systems by balancing these effects would be beneficial and relevant for future cooling technologies.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union Seventh Framework Program (FP7/2007-2013) under Grant Agreement No. 308850 (INFERNOS).

[1] J. Wang, Y. Lai, Z. Ye, J. He, Y. Ma, and Q. Liao, Phys. Rev. E **91**, 050102 (2015).
[2] J.-H. Jiang, O. Entin-Wohlman, and Y. Imry, Phys. Rev. B **85**, 075412 (2012).
[3] D. Venturelli, R. Fazio, and V. Giovannetti, Phys. Rev. Lett. **110**, 256801 (2013).
[4] J. P. Palao, R. Kosloff, and J. M. Gordon, Phys. Rev. E **64**, 056130 (2001).
[5] N. Linden, S. Popescu, and P. Skrzypczyk, Phys. Rev. Lett. **105**, 130401 (2010).
[6] A. Levy and R. Kosloff, Phys. Rev. Lett. **108**, 070604 (2012).
[7] U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012).
[8] R. Kosloff and A. Levy, Annu. Rev. Phys. Chem. **65**, 365 (2014).
[9] B. Cleuren, B. Rutten, and C. Van den Broeck, Phys. Rev. Lett. **108**, 120603 (2012).
[10] A. Levy, R. Alicki, and R. Kosloff, Phys. Rev. Lett. **109**, 248901 (2012).
[11] B. Cleuren, B. Rutten, and C. Van den Broeck, Phys. Rev. Lett. **109**, 248902 (2012).
[12] A. E. Allahverdyan, K. V. Hovhannisyan, and G. Mahler, Phys. Rev. Lett. **109**, 248903 (2012).
[13] O. Entin-Wohlman and Y. Imry, Phys. Rev. Lett. **112**, 048901 (2014).
[14] A. Levy, R. Alicki, and R. Kosloff, Phys. Rev. E **85**, 061126 (2012).
[15] L. E. Reichl and I. Prigogine, *A Modern Course in Statistical Physics* (University of Texas Press, Austin, 1980), Vol. 71, Chap. 7.H.2.
[16] E. M. Loebl, J. Chem. Educ. **37**, 361 (1960).
[17] L. Masanes and J. Oppenheim, arXiv:1412.3828.

II

Influence of measurement error on Maxwell's demon

Vegard Sørdal,[1] Joakim Bergli,[1] and Y. M. Galperin[1,2]
[1]*Department of Physics, University of Oslo, 0316 Oslo, Norway*
[2]*A. F. Ioffe Physico-Technical Institute of Russian Academy of Sciences, 194021 St. Petersburg, Russia*
(Received 17 January 2017; revised manuscript received 29 March 2017; published 21 June 2017)

In any general cycle of measurement, feedback, and erasure, the measurement will reduce the entropy of the system when information about the state is obtained, while erasure, according to Landauer's principle, is accompanied by a corresponding increase in entropy due to the compression of logical and physical phase space. The total process can in principle be fully reversible. A measurement error reduces the information obtained and the entropy decrease in the system. The erasure still gives the same increase in entropy, and the total process is irreversible. Another consequence of measurement error is that a bad feedback is applied, which further increases the entropy production if the proper protocol adapted to the expected error rate is not applied. We consider the effect of measurement error on a realistic single-electron box Szilard engine, and we find the optimal protocol for the cycle as a function of the desired power P and error ε.

DOI: 10.1103/PhysRevE.95.062129

I. INTRODUCTION

Maxwell's demon was introduced as a thought experiment to illustrate the statistical nature of the second law of thermodynamics [1]. The demon has very sharp powers of observation, so it can detect the motion of individual molecules. In addition, it can rapidly act on the basis of its observations and thereby sort fast and slow molecules. This makes heat flow from the cold to the hot side, apparently without the need for any work, in contradiction to the second law of thermodynamics. For some time it was thought that the act of observation necessarily required some amount of work [2,3]. The present consensus [4,5] seems to be that the observation, in principle, can be performed without work. At the same time, the erasure of the information obtained, being a logically irreversible operation, also is thermodynamically irreversible and has a necessary cost in terms of work that is converted to heat. However, there is still some controversy on this point [6–8].

Modern technology now enables us to be as accurate in observation and quick in action as the imagined demon. Recently, several experiments, in which close analogies to the original thought experiment were realized, have been reported in a range of physical systems: atoms [9–11], colloidal particles [12,13], molecules [14], electrons [15–17], and photons [18]. This shift from imagined to real experiments motivates us to study the impact of measurement errors on the performance of experimental Maxwell's demons.

If there is some chance that the measurement result is wrong, it means that the correlation between the state of the system and the measurement device is not perfect. That is, the mutual information between the two is less than the full information of the logical states of the measurement device. In [8], Sagawa and Ueda show that the traditional Landauer bound $W \geqslant T \ln 2$ (we use units where the Boltzmann constant $k_B = 1$) only holds for a symmetric memory, and the total work expended on measurement and memory erasure has a lower bound given by the mutual information I between the system and the measurement device,

$$W_{\text{measure}} + W_{\text{erase}} \geqslant T I. \tag{1}$$

The right-hand side is exactly the same as the heat that can be extracted from a thermal bath using the information about the system. Although measurement errors will give reduced mutual information, we argue that it will not be possible to reach equality in Eq. (1) in this case. To justify this, consider the extreme case in which the mutual information I is zero, i.e., there is a 50% chance that the measurement is wrong. In this case, the measurement can be done reversibly without any work, but there will still be one bit of information stored in the memory that has to be erased with a cost of $T \ln 2$ according to Landauer.

II. ANALYSIS OF A MODEL SYSTEM

To clearly show the difference between a true measurement error and a process that saturates Eq. (1), we will analyze a simple model. Consider a total system (memory + system) with a phase space $\mathcal{P}$. We divide its phase space in subspaces $\mathcal{P}_i$, each of which corresponds to a specific logical state. With the probability distribution of the total phase space denoted $P(x)$, the probability distribution of the logical states is

$$P_L(i) = \sum_{x \in \mathcal{P}_i} P(x) \tag{2}$$

and the conditional probability of the microstate x given the logical state i is

$$P(x|i) = P(x)/P_L(i). \tag{3}$$

The total entropy S, logical entropy (information) H, and conditional entropy $S(\mathcal{P}_i|i)$ are then given by

$$S = -\sum_x P(x) \ln P(x), \quad H = -\sum_i P_L(i) \ln P_L(i),$$

$$S(\mathcal{P}_i|i) = -\sum_{x \in \mathcal{P}_i} P(x|i) \ln P(x|i). \tag{4}$$

The conditional entropy can be thought of as the internal physical entropy of the distribution $P(x|i)$ on $\mathcal{P}_i$ for each of the logical states i. The average conditional entropy is $S_{\text{in}} = \sum_i P_L(i) S(\mathcal{P}_i|i)$, which we call the internal entropy. It follows that we can write the total entropy as a sum,

$$S = H + S_{\text{in}}. \tag{5}$$

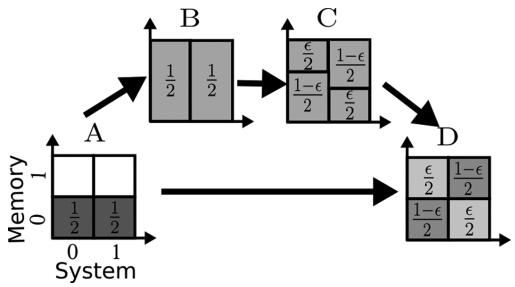

FIG. 1. Schematic of reversible and irreversible measurement in a two-bit total system (system + memory). The four logical subspaces are 00/01/10/11.

With this formalism, we can analyze the model system shown in Fig. 1. The system is a standard Szilard engine, with a single molecule in a box with a dividing wall that can be inserted, removed, and used as a piston. The memory is represented by an equivalent single-molecule box. Consequently, we have four logical states. The phase space of each molecule is reduced to one dimension by only considering the movement of the molecule in the direction in which the volume of the compartments expands and contracts, and ignoring the momentum, as all processes will be isothermal and therefore the momentum distribution is independent of the protocol. The relevant part of the total phase space is then two-dimensional, and we represent the position of the molecule in the system on the horizontal axis, and in the memory on the vertical axis. To calculate the total entropy, we use Eq. (5) and the fact that the conditional entropy of a system uniformly distributed in a given region of phase space is given by the logarithm of the phase-space volume, which we show in the following subsection.

A. Conditional entropy

The free-energy of an ideal gas in a three-dimensional box is

$$F(T,V,N) = -NT \ln\left[\frac{Ve}{N}\left(\frac{mT}{2\pi\hbar^2}\right)^{3/2}\right], \tag{6}$$

which we use to calculate the entropy,

$$S = -\left(\frac{\partial F}{\partial T}\right)_{V,N} = N\left\{\frac{3}{2} + \ln\left[\frac{eV\sqrt{2}}{4N}\left(\frac{mT}{\pi\hbar^2}\right)^{3/2}\right]\right\}$$
$$= N\left[\frac{3}{2} + \ln\left(\frac{V}{NV_q}\right)\right], \quad \frac{1}{V_q(T)} \equiv \frac{e\sqrt{2}}{4}\left(\frac{mT}{\pi\hbar^2}\right)^{3/2}. \tag{7}$$

Here V_q is of the order of the de Broglie wavelength. To keep the classical limit, we have to assume that $V \gg NV_q$. Further, we will deal with one particle. Therefore,

$$S = \frac{3}{2} + \ln\left(\frac{V}{V_q}\right). \tag{8}$$

Our system consists of two volumes (system and memory) with one particle in each, and we need to calculate the conditional entropy for the configurations A and D of Fig. 1 (hereafter denoted $1A$ and $1D$). In each of the four logical states, $i = 00,01,10,11$, the internal states of the system and memory are uncorrelated, and the conditional entropy is the sum of two contributions of the type shown in Eq. (8). Denoting the position of the dividing wall in the system as x_S and in the memory x_M and the length of the box with the gas L, the conditional entropy in the logical state i can be specified as

$$S(\mathcal{P}_i|i) = S_0 + \ln\left(\frac{x_S x_M}{L^2}\right),$$
$$S_0 \equiv 3 + 2\ln\left(\frac{V}{V_q}\right) \gg 1. \tag{9}$$

In the following, we omit the constant S_0, which means that the entropy of the reference state where the dividing walls are removed is zero, and all given entropies are entropy changes from this reference state.

B. Irreversible measurement

We perform a measurement on the system and store it in the memory. Throughout the paper, we assume that the measurement is classical and does not affect the state of the system. If there is a probability ε that the measurement gives the wrong result, we have a transition from A to D in Fig. 1. For configuration $1A$, we have two states with probabilities $\frac{1}{2}$ each, giving

$$H^A = -\sum_i P_L(i)\ln P_L(i) = -2\times\frac{1}{2}\ln\frac{1}{2} = \ln 2. \tag{10}$$

Using Eq. (9) with $x_S = x_M = L/2$ and omitting S_0, we get for the conditional entropy $S(\mathcal{P}_{00}|00) = S(\mathcal{P}_{10}|10) = \ln\frac{1}{4}$ and the internal entropy

$$S^A_{\text{in}} = \sum_i P_L(i) S(\mathcal{P}_i|i) = 2\times\frac{1}{2}\ln\frac{1}{4} = -2\ln 2. \tag{11}$$

Consequently, the total entropy of configuration $1A$ is

$$S^A = H^A + S^A_{\text{in}} = -\ln 2. \tag{12}$$

For configuration $1D$, we obtain in a similar way

$$H^D = -2\times\frac{\varepsilon}{2}\ln\frac{\varepsilon}{2} - 2\times\frac{1-\varepsilon}{2}\ln\frac{1-\varepsilon}{2} = \ln 2 + S_\varepsilon, \tag{13}$$
$$S^D_{\text{in}} = 2\times\frac{\varepsilon}{2}\ln\frac{1}{4} + 2\times\frac{1-\varepsilon}{2}\ln\frac{1}{4} = -2\ln 2, \tag{14}$$
$$S^D = H^D + S^D_{\text{in}} = -\ln 2 + S_\varepsilon, \tag{15}$$

where $S_\varepsilon \equiv -\varepsilon\ln\varepsilon - (1-\varepsilon)\ln(1-\varepsilon)$. The negative-valued entropies are due to the omitted constant S_0. We see that the total entropy in the transition from $1A$ to $1D$ is irreversibly increased by an amount $S^D - S^A = S_\varepsilon$. Since both the system and the memory have equal probabilities of being in their two logical states, the logical information in each is $H^D_{\text{system}} = H^D_{\text{memory}} = \ln 2$. The mutual information between the system and memory is

$$I^D = H^D_{\text{system}} + H^D_{\text{memory}} - H^D = \ln 2 - S_\varepsilon.$$

C. Reversible measurement

The transition from configuration $1A$ to $1D$ can also be achieved reversibly while extracting work if we consider the following steps (this process is also considered in [19]):

$1A \rightarrow 1B$. In the transition from $1A$ to $1B$ we isothermally expand the state 0 of the memory. This allows the particle to expand into the full volume of the memory. In this process, work W is performed by the system, and heat $Q = W$ is taken from the reservoir. The entropy change is

$$\Delta S = W/T = \ln 2$$

with a corresponding entropy decrease in the reservoir.

$1B \rightarrow 1C$. We then perform a measurement on the system, and we reinsert the partition wall in the memory according to the result obtained. There is no error in this measurement, and the correlation between the position of the dividing wall of the memory and the position (left or right) of the gas molecule of the system is perfect. Here ε is just a parameter that describes where we insert the divider in the memory. There is no entropy change.

$1C \rightarrow 1D$. We then compress the divider of the memory isothermally back to the central position. In this process, we have to perform work on the system, but an amount less than the work performed by it in the transition from $1A$ to $1B$. The entropy change is

$$\Delta S = W/T = S_\varepsilon - \ln 2.$$

In our view, this process does not represent a real measurement error, which is irreversible and has an associated entropy production S_ε due to Gibbs or environmental course-graining. The final state of this process ($1D$) is the same as the one obtained when there was a measurement error, but the whole process is thermodynamically reversible, and the reduction of the environmental entropy is exactly the same as the increase of the system entropy. In the process, we have extracted the net work from the thermal bath, so that the work of measurement that enters Eq. (1) is $W_{\text{measure}} = -T S_\varepsilon$, which is negative. Erasing the memory requires $W_{\text{erase}} = T \ln 2$ according to the usual Landauer's principle, which gives

$$W_{\text{measure}} + W_{\text{erase}} = T \ln 2 - T S_\varepsilon = T I^D,$$

which saturates the inequality (1).

D. Origin of the irreversible measurement entropy

To get a deeper understanding of the irreversible nature of a measurement with error, consider Fig. 2. In Fig. 2, step A (hereafter $2A$), we have the same initial state as before. Figure 2, step B (hereafter $2B$) shows the state just after the measurement was performed. Most of the initial states in the phase space are mapped to the correct final region, but a small fraction gets mapped to a different region. This corresponds to the cases in which the result of the measurement does not agree with the actual position of the system molecule. If the system and the measurement device constitute an isolated system during the operation, and no other degrees of freedom are involved, the mapping from $2A$ to $2B$ would be described by a deterministic Hamiltonian evolution in time. Liouville's theorem then guarantees that the entropy of the final state is the same as that in the initial state. If the evolution is affected by other microscopic degrees of freedom in the device or the environment, which is certainly realistic in most cases, the mapping would be stochastic, and it depends on these additional degrees of freedom. We assume that after $2B$, the phase points will never again cross the lines separating the different logical states. The physical meaning of this is that the barriers between the states are infinite. In a short time, the phase-space region where the system can be found will develop into some complicated shape $2C$, but for a closed system the entropy will still be the same. Now we have to appeal to some coarse-graining procedure. For a closed system, we refer to the phase-space coarse-graining introduced by Gibbs (see Ref. [20] for a recent discussion). In the presence of some interaction with an environment, coarse-graining occurs over dynamical evolution [21,22]. In this way, the complex structure of the accessible phase space in $2C$ is rendered indistinguishable and is replaced by the uniform distribution in $2A$. This step is irreversible and increases the total entropy of the system by S_ε without any decrease in entropy anywhere.

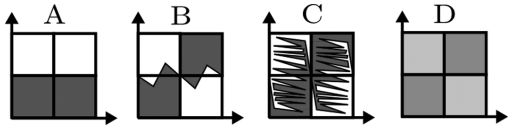

FIG. 2. How a system evolves from step A to D in Fig. 1 after a measurement error.

III. MODEL FOR A SZILARD ENGINE WITH MEASUREMENT ERRORS

To study the effects of this measurement error, we will now analyze a model of an experimentally realized Szilard engine [15]. This model is comprised of a single-electron box consisting of two metallic islands connected by a tunnel junction. The existence of an additional electron on one of the two islands can be measured by the charge configuration of the box, and its state can be controlled by gate voltages applied to the islands, giving a time-dependent potential difference $V(t)$ between the two islands. Work can be extracted from the system via the following procedure:

(i) Perform a measurement, and quickly set the potential of the occupied island to zero while raising the potential of the empty island to some value $V_0 \equiv V(0^+)$.

(ii) Reduce the potential of the raised island according to some protocol $V(t)$ until time $t = \tau$, at which point we start over from step (i).

There is a probability that the electron will tunnel between the two islands, and whenever the electron occupies the island where the potential is being decreased, heat is extracted from the environment and converted to work. We imagine that we are continuously repeating steps (i) and (ii) above, and we want to minimize the total entropy production rate when varying the driving protocol $V(t)$ and the time τ. In the experimental protocol in [15], the potential difference was always reduced back to zero before the next measurement. In our optimal protocol, it does not need to be zero since the energy gained from tunneling decreases with the potential difference, and longer τ gives smaller power.

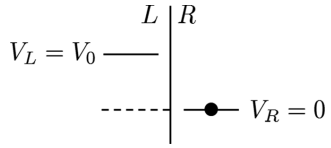

FIG. 3. Level diagram after the first measurement and raising the potential of the empty island that does not cost any work.

Details of the protocol

We have a set of two islands—left (L) and right (R)—divided by a tunnel barrier. The islands are gated, so their potentials can be manipulated independently. The pair of islands contains only one excess electron, so each island may contain either zero or one excess electron; the occupancy of each island can be measured, say, by single electron transistor. A similar setup was used in the experiment [15]. Let us start the protocol from the state where we have measured an electron on the right island. Then we quickly raise the potential of the empty island to the value V_0. After this procedure, we arrive at the level diagram shown in Fig. 3. Then we decrease the potential of the raised island according to some protocol $V(t)$ in the time $t = [0,\tau]$, as shown in Fig. 4 (left panel). At time $t = \tau$, we reach the situation shown in the right panel of Fig. 4, where the electron can be found on the left island with probability p_τ and on the right island with probability $(1 - p_\tau)$. If we measure the system at that moment, we can arrive at the level diagram shown in Fig. 5(a) (with probability p_τ), or at the diagram shown in Fig. 5(b) [with probability $(1 - p_\tau)$]. In the first case, we quickly decrease the energy of the left island extracting the work V_τ, and we raise the energy of the left island to the value V_0 (with no cost). The average extracted work is then $\langle W\rangle = p_\tau V_\tau$. Then we arrive at the diagram shown in Fig. 5(c), which is a mirror of Fig. 3. It is therefore thermodynamically equivalent and we have a completed cycle. In the case shown in Fig. 5(b), we quickly move the left level up to V_0 (with no cost) arriving at the situation shown in Fig. 3, again completing the cycle. There exists another protocol leading to the same consequences. Namely, at the time instance $t = \tau$ one can quickly decrease the potential of the left island to zero (before the measurement). The extracted work is V_τ while the probability that the left island was occupied is p_τ, so the average work is $\langle W\rangle = p_\tau V_\tau$. Then we measure the position of the electron and raise the level of the empty island up to the value V_0. Again, the cycle is closed.

IV. DERIVING THE OPTIMAL PROTOCOL

A thermodynamically equivalent turnstile version of this model was previously analyzed [23] when there were no errors in the measurements, and the consequences of reduced mutual

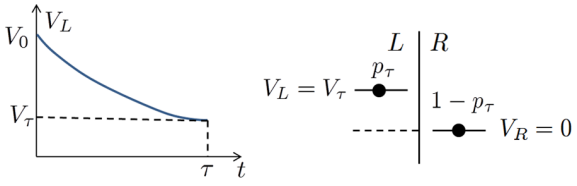

FIG. 4. Left: Example protocol during time $0 < t < \tau$. Right: Level diagram at $t = \tau$.

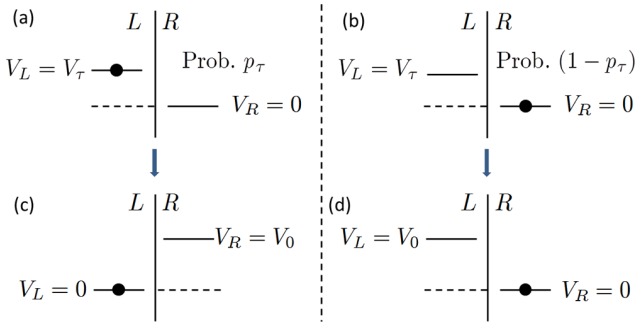

FIG. 5. Completing the cycle.

information were discussed [24]. If there is an error in the measurement, we have an additional entropy production term S_ε to the total entropy production, and the feedback operation $V(t)$ will have to be adapted to the expected error to minimize the entropy ΔS that is produced during feedback operation. The total entropy produced in a cycle is then given by

$$\Delta S_{\rm tot} = \underbrace{S_\epsilon}_{\text{measurement}} + \underbrace{\Delta S - Q/T}_{\text{operation}}, \tag{16}$$

where Q is the heat exchanged between the system and the environment. Extending the analysis from [23] to finite error is principally not difficult. We minimize the total entropy production rate,

$$\dot S_{\rm tot} = \frac{\Delta S_{\rm tot}}{\tau} = \frac{S_\varepsilon}{\tau} + \frac{\Delta H}{\tau} - P, \tag{17}$$

using the Euler-Lagrange formalism, which leads to a nonlinear differential equation that has to be solved numerically. Here $\Delta S = \Delta H$ since we assume there are no excitations from the ground state of the islands, therefore the internal entropy change is $\Delta S_{\rm in} = 0$. $P = Q/(T\tau)$ is the rate of heat exchange between the system and the environment.

A. Thermodynamic properties

Let $p_1(t)$ and $p_2(t)$ be the probabilities to find the system in state 1 (the right island) and 2 (the left island), respectively. The transitions between these two states are described by the rates Γ_{12} and Γ_{21}, which satisfy the detailed balance $\Gamma_{21}/\Gamma_{12} = e^{\Delta E/T}$, where ΔE is the difference in energy of the two states (note that since ΔE is a function of time, the rates will also be time-dependent). The master equations are thus

$$\begin{aligned} \dot p_1 &= -\Gamma_{12} p_1 + \Gamma_{21} p_2 = -\Gamma p_1 + \Gamma_{21}, \\ \dot p_2 &= \quad \Gamma_{12} p_1 - \Gamma_{21} p_2 = -\Gamma p_2 + \Gamma_{12}, \end{aligned} \tag{18}$$

where $\Gamma(t) \equiv \Gamma_{12}(t) + \Gamma_{21}(t)$. As in Refs. [23,25], for the sake of simplicity we choose Γ to be independent of time, and we set $\Gamma = 1$. The energy of state i is denoted $E_i(t)$, and in the protocol described above we have $E_1(t) = 0$ and $E_2(t) = V(t)$. The total work extracted during the period τ is

$$W_{\rm ex} = -\sum_{i=1}^{2} \int_0^\tau dt\, p_i \dot E_i, \tag{19}$$

the change in internal energy of the system is

$$\Delta U = \sum_{i=1}^{2} [p_i(\tau)E_i(\tau) - p_i(0)E_i(0)], \tag{20}$$

and the transferred heat from the environment to the system is

$$Q = \Delta U + W_{\rm ex} = \sum_{i=1}^{2} \int_0^\tau dt\, \dot{p}_i E_i(t). \tag{21}$$

The information entropy associated with the measurement is $H = -\sum_{i=1}^{2} p_i \ln p_i$, and the change in information entropy can be written as an integral,

$$\begin{aligned}\Delta H &= H_\tau - H_0 \\ &= -\sum_{i=1}^{2} [p_i(\tau)\ln p_i(\tau) - p_i(0)\ln p_i(0)] \\ &= -\sum_{i=1}^{2} \int_0^\tau dt\, \dot{p}_i(t)\ln p_i(t).\end{aligned} \tag{22}$$

Since $p_1(t) = 1 - p_2(t)$, we can relabel $p_2(t) \equiv p$, and we write the entropy produced per cycle as

$$\frac{\Delta H}{\tau} = -\frac{1}{\tau}\int_0^\tau dt\, \dot{p} \ln\left(\frac{p}{1-p}\right). \tag{23}$$

From the master equation (18), we get

$$\dot{p} = -p + \frac{1}{e^V + 1}, \tag{24}$$

where from now on we relabel $V(t) \equiv V$, and we measure time in units of Γ and energy in units of T. From this equation, we can express

$$V = \ln\left(\frac{1}{p+\dot{p}} - 1\right).$$

The power is defined as the average heat extracted from the reservoir per cycle τ, $P = Q/(T\tau)$, and it can be written as

$$P = \frac{1}{\tau}\int_0^\tau dt\, \dot{p}V = \frac{1}{\tau}\int_0^\tau dt\, \dot{p}\ln\left(\frac{1}{p+\dot{p}} - 1\right). \tag{25}$$

We are interested in the optimal protocol for the measurement and erasure cycle. In this system, the optimal protocol means finding the protocol $V(t)$ and the total time τ, which minimize the entropy production rate at a given measurement error ε. We also set the value of the power P, given by Eq. (25), to see how the solutions depend on the power we want to extract. The initial condition is set by $p(t=0) = \varepsilon$. That is, there is a chance, ε, that the electron was on the opposite island of what we measured.

B. Minimizing the entropy production rate

Since we want to minimize the entropy production rate while keeping the power at a finite value P, we have to introduce the Lagrange multiplier λ to obtain the functional

$$J = \frac{S_\epsilon}{\tau} + \frac{\Delta H}{\tau} + \lambda P = \frac{S_\epsilon}{\tau} + \frac{1}{\tau}\int_0^\tau dt\, L(p,\dot{p},\lambda), \tag{26}$$

with the Lagrangian

$$L(p,\dot{p},\lambda) = \left[-\ln\left(\frac{p}{1-p}\right) + \lambda \ln\left(\frac{1}{\dot{p}+p} - 1\right)\right]\dot{p}. \tag{27}$$

Using the Euler-Lagrange equation

$$\frac{\partial L}{\partial p} = \frac{d}{dt}\frac{\partial L}{\partial \dot{p}}, \tag{28}$$

we obtain the following second-order nonlinear ordinary differential equation:

$$\ddot{p} = \frac{\dot{p}^2(\dot{p} + p - 1/2)}{p(\dot{p}+p-1) + \dot{p}/2}. \tag{29}$$

To solve this equation, we need to impose a set of constraints to the solutions we want. The first constraint is that the power has to be a finite fixed value P, given by Eq. (25):

$$G(\tau,p,\dot{p}) \equiv P - \frac{1}{\tau}\int_0^\tau dt\, \dot{p}\ln\left(\frac{1}{p+\dot{p}} - 1\right) = 0. \tag{30}$$

The second constraint comes from a consideration of the end-point values of $p(t)$. The initial condition of $p(t)$ is given by $p(0) = \varepsilon$, but since the value of $p(t)$ is not fixed at the end point $p(\tau) = p_\tau$, we have a second constraint, $(\partial L/\partial \dot{p})_{t=\tau} = 0$, which can be written as

$$\begin{aligned}F_1(\lambda,\tau,p) &\equiv \lambda\left[\ln\left(\frac{1}{p_\tau + \dot{p}_\tau} - 1\right) + \frac{\dot{p}_\tau}{(\dot{p}_\tau + p_\tau - 1)(\dot{p}_\tau + p_\tau)}\right] \\ &\quad - \ln\left(\frac{p_\tau}{1-p_\tau}\right) = 0.\end{aligned} \tag{31}$$

The third and final constraint is due to the fact that variation of Eq. (17) with respect to the period τ should be zero. It is given by

$$\frac{\partial J}{\partial \tau} = \lambda\frac{\partial P}{\partial \tau} + \frac{\partial}{\partial \tau}\frac{\Delta H}{\tau} - \frac{S_\varepsilon}{\tau^2} = 0, \tag{32}$$

where

$$\begin{aligned}\frac{\partial P}{\partial \tau} &= \frac{\partial}{\partial \tau}\left[\frac{1}{\tau}\int_0^\tau dt\, \dot{p}\ln\left(\frac{1}{p+\dot{p}} - 1\right)\right] \\ &= \frac{\dot{p}_\tau}{\tau}\ln\left(\frac{1}{p_\tau + \dot{p}_\tau} - 1\right) - \frac{P}{\tau}\end{aligned} \tag{33}$$

and

$$\begin{aligned}\frac{\partial}{\partial \tau}\frac{\Delta H}{\tau} &= \frac{\partial}{\partial \tau}\left[-\frac{1}{\tau}\int_0^\tau dt\, \dot{p}\ln\left(\frac{p}{1-p}\right)\right] \\ &= -\frac{1}{\tau}\dot{p}_\tau \ln\left(\frac{p_\tau}{1-p_\tau}\right) - \frac{\Delta H}{\tau^2}.\end{aligned} \tag{34}$$

The full equation for the third constraint is thus

$$\begin{aligned}F_2(\lambda,\tau,p) \equiv &\left[\ln\left(\frac{1-p_\tau}{p_\tau}\right) + \lambda\dot{p}_\tau \ln\left(\frac{1}{p_\tau + \dot{p}_\tau} - 1\right)\right] \\ &-\lambda P - \frac{1}{\tau}[\Delta H + S_\varepsilon] = 0.\end{aligned} \tag{35}$$

This constraint can be combined with the free-end-point constraint by eliminating the Lagrange multiplier λ

to obtain

$$F(\tau, p_\tau, \dot{p}_\tau) \equiv \ln\left(\frac{p_\tau}{1-p_\tau}\right)\left[P(\dot{p}_\tau + p_\tau)(p_\tau + \dot{p}_\tau - 1) + \dot{p}_\tau^2\right] - \frac{S_\tau}{\tau}\left[\dot{p}_\tau + \ln\left(\frac{1}{p_\tau + \dot{p}_\tau} - 1\right) \times (\dot{p}_\tau + p_\tau)(p_\tau + \dot{p}_\tau - 1)\right] = 0, \tag{36}$$

where $S_\tau = \Delta H + S_\varepsilon = -p_\tau \ln p_\tau - (1-p_\tau)\ln(1-p_\tau)$ is the entropy of the system at time $t = \tau$. Euler's method is then used to solve the second-order differential equation in Eq. (29). Since it is a second-order equation, we have two constants that need to be fixed (τ and V_0). We find these values as the roots of the two constraints in Eqs. (30) and (36) by using Newton's method.

V. RESULTS

We now present the main results of this analysis. The model has a parameter Γ that determines the tunneling rate between the two islands, and we measure time in units of Γ^{-1} and energy in units of temperature T. In Fig. 6 we plot the optimal period τ as a function of the power P for selected values of the error ε.

A. High power limit

We find that there is a maximal amount of power one can extract, $P^{\max}(\varepsilon)$. As this value is reached, τ approaches zero linearly, and the entropy production rate diverges to infinity. To leading order we have

$$\tau \propto P^{\max} - P, \tag{37}$$

$$\dot{S}_{\text{tot}} \propto \frac{1}{P^{\max} - P}. \tag{38}$$

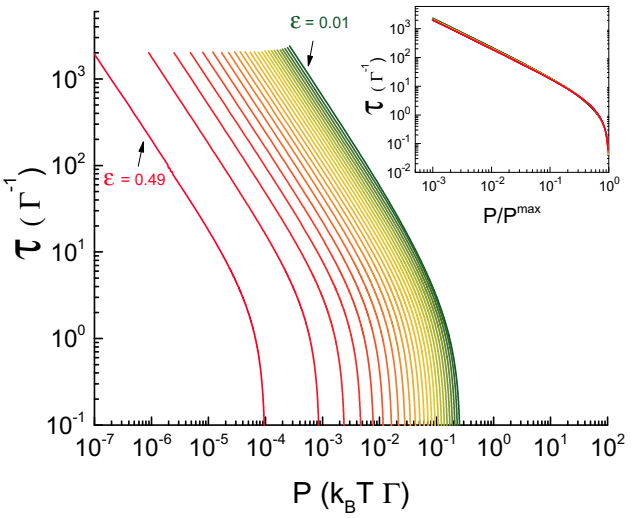

FIG. 6. The main figure shows τ as a function of P for different ε (in steps of 0.02). The inset gives the scaled form of the same data, with τ as a function of $P/P^{\max}$.

If τ is plotted as a function of $P/P^{\max}$, the scaled graphs are close to collapsing over the whole range of powers, as shown in the inset of Fig. 6. We can always find the value of $P^{\max}$ numerically, but we can also derive a single transcendental equation that determines $P^{\max}$, and in the case of error-free measurements we can also solve it analytically. By taking the limit as $\tau \to 0$ in Eq. (25), we find that

$$P^{\max} = V_0 \dot{p}_0 = V_0\left(\frac{1}{e^{V_0}+1} - \varepsilon\right), \tag{39}$$

which expresses $P^{\max}$ in terms of V_0. Consider Eq. (36) when $\tau \to 0$. Since the other terms are finite, the only way to avoid a divergence of the last term is for the expression in the large square brackets to be zero. For $\tau = 0$ we have $p_\tau = \varepsilon$, and with Eq. (24) we find that V_0 satisfies the equation

$$1 + (1 - V_0)e^{V_0} - \varepsilon(e^{V_0}+1)^2 = 0. \tag{40}$$

For $\varepsilon = 0$ we find that the maximum power is given by the Lambert W function,

$$P^{\max} = W(e^{-1}) = 0.278\,46\ldots \tag{41}$$

with the initial value of the potential $V_0 = 1 + W(e^{-1})$. This analytical result is in perfect agreement with our numerical result.

B. Low power limit

When $P = 0$, we can assume that the system is always in equilibrium at the given value of V, which means that $p = p_a = (e^V + 1)^{-1}$. We assume for small P that we have $p = p_a + O(P)$, and that $\tau = A/P$. Inserting into Eq. (25) and expanding in P, we find that it becomes

$$1 = \frac{1}{A}\int_0^\infty dt\, V\dot{p}_a + O(P). \tag{42}$$

In the limit $\tau \to \infty$, corresponding to quasistatic operation, the entropy production will vanish if

$$(e^{V_0}+1)^{-1} = p_0 = \varepsilon \to V_0 = \ln\left(\frac{1}{\varepsilon} - 1\right), \tag{43}$$

as shown in [24]. It is reasonable, and confirmed by the numerical solution, that at small P and long time τ the potential difference will be brought all the way back to zero, $V(\tau \to \infty) = 0$. With these two boundary values, we get

$$A = \int_0^\infty dt\, V\dot{p}_a = -\int_0^{V_0} V dV \frac{dp_a}{dV} = \ln 2 - S_\varepsilon. \tag{44}$$

The end result is that in the limit of low power, $P \to 0$, the optimal period τ diverges as

$$\tau = (\ln 2 - S_\varepsilon)P^{-1}, \tag{45}$$

in agreement with our numerical result. In the polynomial expansion as $P \to 0$ of the total entropy production, $\Delta S_{\text{tot}} = c_0 + c_1 P$, we know that for perfect measurements c_0 has to be zero since there is no entropy production during reversible operation. We have $\tau = (\ln 2 - S_\varepsilon)P^{-1}$ for small P, and therefore we get $\dot{S}_{\text{tot}} = \Delta S_{\text{tot}}/\tau \propto P^{-2}$, in agreement with [23]. If errors are present, the measurement entropy S_ε exists

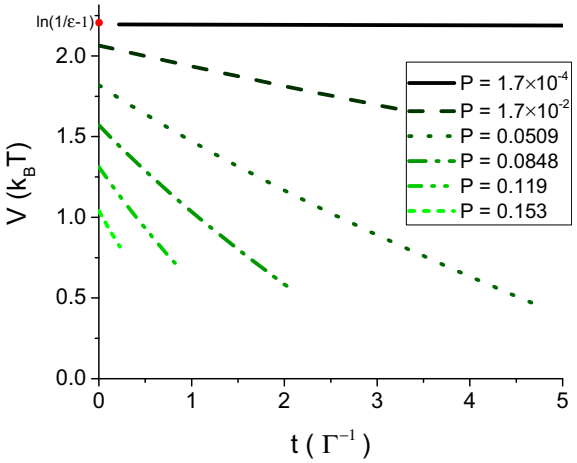

FIG. 7. $V(t)$ for $\varepsilon = 0.1$ and several values of P.

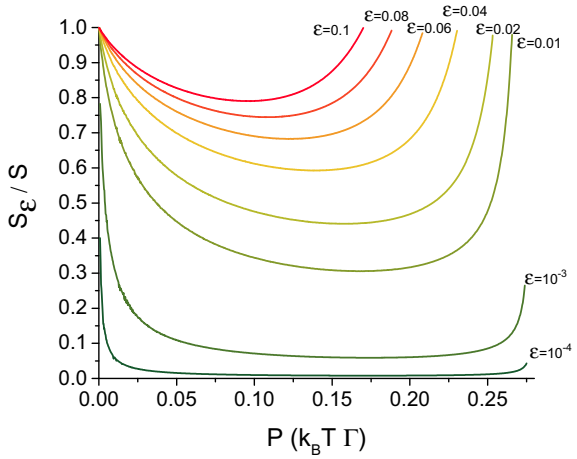

FIG. 8. The fraction of the total entropy production $S_{\rm tot}$ that is due to the measurement error S_ε as a function of power for various values of ε.

even for reversible operation so that $c_0 = S_\varepsilon$, and we obtain an additional linear behavior of the entropy production rate

$$\dot{S}_{\rm tot} = S_\varepsilon(\ln 2 - S_\varepsilon)^{-1} P + c_1^* P^2, \tag{46}$$

which we confirm numerically in Appendix 2.

C. The optimal protocol

Examples of optimal protocols for $\varepsilon = 0.1$ and several values of P are shown in Fig. 7. We observe that the time τ before the protocol should be repeated decreases with increasing P, and the initial value V_0 increases with decreasing P. The quasistatic limit ($\tau \to \infty / P \to 0$) found in Eq. (43) was $V_0 = \ln(\frac{1}{\varepsilon} - 1)$, which is marked in the plot. See Appendix 1 for more details on the behavior of $V(t)$ and $p(t)$ at time τ.

To extract maximum power, one has to balance the following: the amount of energy gained per tunneling event, the probability that tunneling occurs, and the probability of back-tunneling while reducing the potential difference. These results tell us that the maximum power is reached with rapid measurements, favoring low probability high-energy tunneling events, and a steeply sloped $V(t)$. However, this comes at the cost of divergence in the entropy production rate. This result is obtained assuming a constant total tunneling rate Γ, and it may change for systems in which Γ depends on the potential difference between the two islands.

D. Effect of the measurement error

To clearly see the effect of the measurement error on the total entropy production, we plot in Fig. 8 the ratio $S_\varepsilon / S_{\rm tot}$ for various values of ε. For $P \to 0$ we approach reversible operation ($\Delta S = 0$), and all of the total entropy production is due to the measurement error. When $P \to P^{\rm max}$, the measurement entropy dominates again since there is no time for heat transfer from the environment when $\tau \to 0$. For vanishingly small errors, its effect is only noticeable at the boundary values of P, but even for relatively minor measurement errors a significant portion of the entropy production is due to the measurement error for all P.

VI. SUMMARY

If we make an error in a measurement, there is an associated net entropy production. This applies to measurements of any type and with an arbitrary number of outcomes. For a symmetric binary measurement where the probability of error is ε, the entropy increases by the amount S_ε. This entropy increase can be understood from a coarse-graining of either the phase space (for a closed system) or the dynamical evolutions (for an open system). We have investigated the consequences of a finite error on the optimal performance of a realistic Szilard engine at finite (given) power. We found the existence of a maximal power $P^{\rm max}$, which also exists for error-free measurements, and which decreases with increasing error. The entropy production rate diverges as the maximal power is approached. For small power, the entropy production rate is quadratic in P in the absence of errors, but it becomes linear when errors are present.

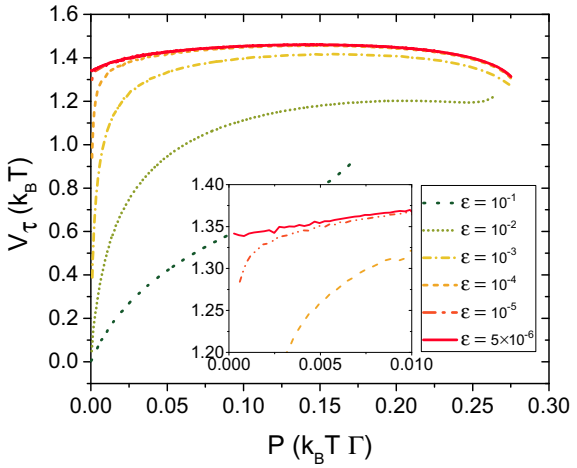

FIG. 9. V_τ as function of P for different ε. The inset shows enlarged what happens for small errors and powers.

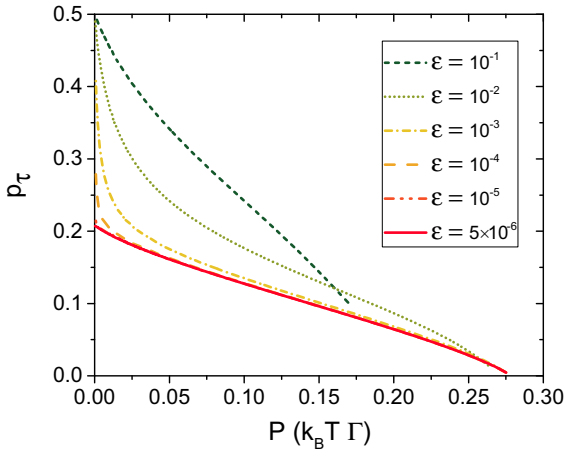

FIG. 10. p_τ as a function of P for different ε.

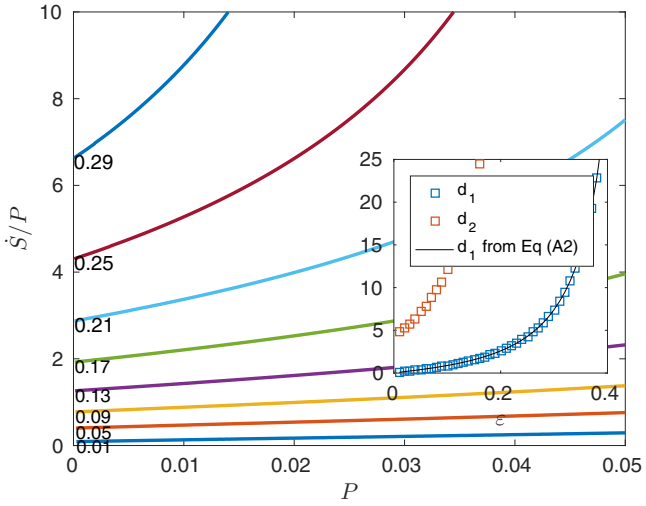

FIG. 11. $\dot{S}/P$ as a function of P with labels on the curves giving ε. For each curve, the value at $P = 0$ and the slope of the tangent at that point will give the coefficients c_1 and c_2 of Eq. (A1). These are shown as functions of ε in the inset, together with c_1 from Eq. (A2).

ACKNOWLEDGMENT

We are grateful to J. Pekola and I. Khaymovich for illuminating discussions.

APPENDIX: ADDITIONAL RESULTS

Here we present some additional results on the optimal protocol.

1. $V(t)$ and $p(t)$ at time τ

Figure 9 shows $V(\tau) = V_\tau$ as a function of P for different ε. While it seems that for any finite ε we find $V_\tau \to 0$ as $P \to 0$, we see that for small ε one has to go to very small powers to see this, and for most powers V_τ is between 1 and 1.5. This indicates a singular behavior of the function $V_\tau(P,\varepsilon)$ at $P = 0$ and $\varepsilon = 0$, and the limiting value will depend on how this point is approached. In Ref. [23] we found that $V_\tau = 1.33$ for $\varepsilon = 0$ and small P. From Fig. 9 (inset) we can see that this agrees well with what we would expect if we first took the limit $\varepsilon \to 0$ and then $P \to 0$. The same singularity is reflected in the probability p_τ to find the electron on the opposite island at time τ from the one it was measured at time 0 as shown in Fig. 10. For all finite ε we have $\lim_{P\to 0} p_\tau = 0.5$, but for small ε this only happens at very small P.

2. Polynomial expansion of $\dot{S}$

In Ref. [23] it was found that for $\varepsilon = 0$ and small P, $\dot{S}$ is proportional to P^2. We find that this is not true for finite ε. We expand to second order,

$$\dot{S} = d_1 P + d_2 P^2, \tag{A1}$$

where d_1 and d_2 are functions of ε. Plotting $\dot{S}/P$ as a function of P (Fig. 11), we get d_1 and d_2 as the intercept and slope of the tangent at $P \to 0$ (Fig. 11, inset). From the analytical results in Eq. (46), we know that $d_1 = S_\varepsilon(\ln 2 - S_\varepsilon)^{-1}$,

$$\dot{S} = d_1 P \quad \text{with} \quad d_1 = S_\varepsilon(\ln 2 - S_\varepsilon)^{-1}, \tag{A2}$$

which, as shown in Fig. 11 (inset), agrees perfectly with the numerical solution.

[1] *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*, edited by H. S. Leff and A. Rex (Institute of Physics, Bristol, 2003).

[2] L. Szilard, Über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen, Z. Phys. **53**, 840 (1929).

[3] L. Brillouin, *Science and Information Theory* (Academic Press, New York, 1956).

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

[5] C. H. Bennett, The thermodynamics of computation—A review, Int. J. Theor. Phys. **21**, 905 (1982).

[6] J. D. Norton, Waiting for Landauer, Stud. Hist. Philos. Sci. Pt. B **42**, 184 (2011).

[7] J. Ladyman and K. Robertson, Landauer defended: Reply to Norton, Stud. Hist. Philos. Sci. Pt. B **44**, 263 (2013).

[8] T. Sagawa and M. Ueda, Minimal Energy Cost for Thermodynamic Information Processing: Measurement and Information Erasure, Phys. Rev. Lett. **102**, 250602 (2009).

[9] G. N. Price, S. T. Bannerman, K. Viering, E. Narevicius, and M. G. Raizen, Single-Photon Atomic Cooling, Phys. Rev. Lett. **100**, 093004 (2008).

[10] J. J. Thorn, E. A. Schoene, T. Li, and D. A. Steck, Experimental Realization of an Optical One-Way Barrier for Neutral Atoms, Phys. Rev. Lett. **100**, 240407 (2008).

[11] M. G. Raizen, Comprehensive control of atomic motion, Science **324**, 1403 (2009).

[12] 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).

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

[14] V. Serreli, C.-F. Lee, E. R. Kay, and D. A Leigh, A molecular information ratchet, Nature (London) **445**, 523 (2007).

[15] J. V. Koski, V. F. Maisi, J. P. Pekola, and D. V. Averin, Experimental realization of a Szilard engine with a single electron, Proc. Natl. Acad. Sci. USA **111**, 13786 (2014).

[16] J. V. Koski, A. Kutvonen, I. M. Khaymovich, T. Ala-Nissila, and J. P. Pekola, On-Chip Maxwell's Demon as an Information-Powered Refrigerator, Phys. Rev. Lett. **115**, 260602 (2015).

[17] K. Chida, K. Nishiguchi, G. Yamahata, H. Tanaka, and A. Fujiwara, Thermal-noise suppression in nano-scale Si field-effect transistors by feedback control based on single-electron detection, Appl. Phys. Lett. **107**, 073110 (2015).

[18] M. D. Vidrighin, O. Dahlsten, M. Barbieri, M. S. Kim, V. Vedral, and I. A. Walmsley, Photonic Maxwell's Demon, Phys. Rev. Lett. **116**, 050401 (2016).

[19] T. Sagawa and M. Ueda, Role of mutual information in entropy production under information exchanges, New J. Phys. **15**, 125012 (2013).

[20] K. Ridderbos, The coarse-graining approach to statistical mechanics: How blissful is our ignorance? Stud. Hist. Philos. Sci. Pt. B **33**, 65 (2002).

[21] J. M. Blatt, An alternative approach to the ergodic problem, Prog. Theor. Phys. **22**, 745 (1959).

[22] S. Lloyd, Use of mutual information to decrease entropy: Implications for the second law of thermodynamics, Phys. Rev. A **39**, 5378 (1989).

[23] J. Bergli, Y. M. Galperin, and N. B. Kopnin, Information flow and optimal protocol for a Maxwell-demon single-electron pump, Phys. Rev. E **88**, 062139 (2013).

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

[25] D. Mandal and C. Jarzynski, Work and information processing in a solvable model of Maxwell's demon, Proc. Natl. Acad. Sci. USA **109**, 11641 (2012).

III

Quantum particle in a split box: Excitations to the ground state

Vegard B. Sørdal and Joakim Bergli
Department of Physics, University of Oslo, 0316 Oslo, Norway

(Received 11 December 2018; published 20 February 2019)

We discuss two different approaches for splitting the wave function of a single-particle box (SPB) into two equal parts. Adiabatic insertion of a barrier in the center of a SPB in order to make two compartments which each have probability $1/2$ of finding the particle in it is one of the key steps for a Szilard engine. However, any asymmetry between the volume of the compartments due to an off-center insertion of the barrier results in a particle that is fully localized in the larger compartment, in the adiabatic limit. We show that rather than exactly splitting the eigenfunctions in half by a symmetric barrier, one can use a nonadiabatic insertion of an asymmetric barrier to induce excitations to only the first excited state of the full box. As the barrier strength goes to infinity the excited state of the full box becomes the ground state of one of the new boxes. Thus, we can achieve close to exact splitting of the probability between the two compartments using the more realistic nonadiabatic, not perfectly centered barrier, rather than the idealized adiabatic and central barrier normally assumed.

DOI: 10.1103/PhysRevA.99.022121

I. INTRODUCTION

The Szilard engine is a simple conceptual model of an information processing system [1]. The classical model is a single particle in a box, coupled to a thermal bath. By inserting a movable barrier in the center of the box, the probability of finding the particle in either compartment becomes $1/2$. If we now perform a measurement to find out which compartment the particle is in, we generate one bit of Shannon information which is stored in some memory. Since the box is coupled to a thermal bath, we can extract work by allowing the compartment in which we find the particle to expand and fill the whole box. The maximum work extracted in this way is $k_BT \ln 2$, and this is achieved by isothermal expansion. To complete the cycle the memory is deleted, which has a minimum energy cost of $k_BT \ln 2$, according to Landauer's principle [2]. Therefore, if we perform reversible operations, the full cycle of measurement, work extraction, and information deletion generates no entropy. The quantum mechanical version of the Szilard engine is similar, only now we are splitting the wave function of the particle. The quantum measurement and, assuming the memory is classical, deletion is similar to the classical case, but there are subtle differences when it comes to the insertion, expansion, and removal of the barrier [3].

The adiabatic theorem in quantum mechanics tells us that a system remains in its instantaneous eigenstate as long as it has a gapped energy spectrum and the perturbation acting on it is slow enough to prevent transition between the eigenstates. Based on this, it has been remarked in [4] that if the particle is in the ground state and the barrier is inserted off-center, such that one compartment is larger than the other, the particle will always be localized in the larger compartment. This is because the energy spectrum is proportional to L^{-2}, where L is the length of the compartment. The result is independent of how small the asymmetry between the compartments is; any finite difference between the compartment sizes will give the same result.

With modern technology we can now experimentally realize what was before only a thought experiment. In the last decade, the creation of Szilard engines has been reported in a range of physical systems: atoms [5–7], colloidal particles [8,9], molecules [10], electrons [11–13], and photons [14]. In experiments the barrier is not inserted adiabatically nor exactly in the center, and one can ask the question of how the result of the previous paragraph changes when the barrier is inserted at a finite rate.

Although a finite rate of insertion can make the probability of finding the particle in the smaller compartment nonzero, the downside is that a fast rate results in excitations to higher energy levels. The Szilard engine measurement procedure traditionally only determines which side of the box the particle is found, not its exact eigenstate. Therefore excitation of high energy states introduces additional entropy that is not accounted for in the which-side measurement. Information is therefore lost when performing the measurement, leading to decreased efficiency of the Szilard engine.

Previous work [15] has studied the asymmetric insertion of a δ potential barrier at high rates of insertion. In contrast to their work, we expand the full wave function in its instantaneous energy eigenstates and point out the fact that it is possible to asymmetrically insert a barrier with a finite rate and obtain very close to an equal probability distribution, with negligible excitations to higher states than to the first excited state. Therefore, the asymmetric Szilard engine with finite rate of insertion of the barrier can have the same efficiency as the symmetric Szilard engine with adiabatic insertion.

There are two fundamentally different ways to get an equal probability of occupying the left and right box of a Szilard engine. One way is to follow the usual protocol of splitting a symmetric wave function into two exactly equal parts, i.e., inserting a barrier in the center of a box with a particle in the ground state. Figure 1 shows the time evolution of the eigenstates and eigenenergies when inserting a barrier with time-dependent strength $\alpha(t)$ (dashed vertical line), in the

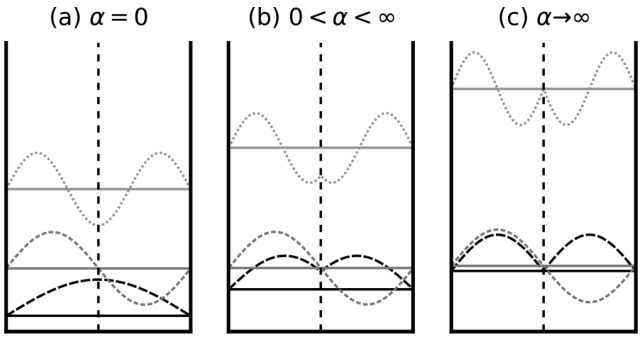

FIG. 1. Schematic of the three first eigenfunction and energies for a symmetric box for three different values of the barrier strength $\alpha(t)$. (a) Initial state of the system, before the barrier is inserted. (b) At an intermediate time before $\alpha(t) \to \infty$. We see that when the barrier is inserted at the center of the box it hits the nodes of the antisymmetric eigenfunctions, and therefore there are no excitations to this state [see Eq. (6)]. (c) The limit when $\alpha(t) \to \infty$. The total wave function is symmetric about the barrier, and the probability of finding the particle in either compartment is $1/2$.

center of the box. Figure 1(a) is the initial state of the system, before the barrier has begun to be inserted. Figure 1(b) is an intermediate step with $0 < \alpha < \infty$ before the two compartments have been completely isolated from each other in Fig. 1(c) as $\alpha \to \infty$. The eigenstates in Fig. 1(c) are split exactly in half, with a probability of $1/2$ on either side.

The second way is to insert the barrier asymmetrically and nonadiabatically, in such a way that only the first excited state is excited; the eigenfunction of the ground state will be large in the larger compartment and small in the smaller compartment, and vice versa for the first excited state. This method is illustrated in Fig 2. The initial state in Fig. 2(a), before the barrier is inserted, is identical to that in Fig. 1(a). However, as the barrier is increased via 2(b) through 2(c) the symmetric eigenfunction becomes zero in the smaller compartment, while the antisymmetric becomes zero in the larger compartment. Of course it has to be this way, since when

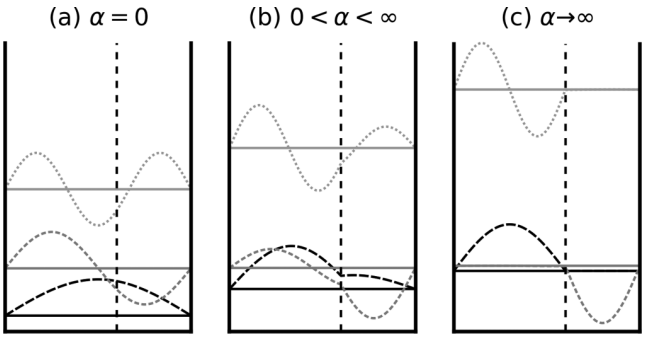

FIG. 2. Same as Fig. 1(a), but for an asymmetric box. (a) The initial state of the system, identical to Fig. 1(a). Only now the eigenfunction of the first excited state is nonzero at the point we insert the barrier, allowing excitations from the ground state. From the intermediate time step in (b) to the final state in (c) the eigenfunction of the ground and first excited states evolves such that it is approximately zero in the smaller and larger compartments, respectively.

$\alpha \to \infty$ what we have is essentially two rescaled copies of the initial state. The third energy level in Fig. 2(a) becomes the new first excited state of the larger compartment in Fig. 2(c), while the first exited state in Fig. 2(a) becomes the new ground state of the smaller compartment in Fig. 2(c). Only exciting the first excited state of the original box still results in no excitations after the measurement, since it becomes the new ground state of the compartment.

A good thought experiment is never set in some complicated system with many degrees of freedom. Rather, it is a surprising result or counterintuitive implication obtained from the study of a simplified model of reality. One might ask why further study of a thought experiment that has already been experimentally realized is necessary. In our opinion there are two main reasons: The first reason is that studying all the aspects of this conceptual model helps us to understand the key physical effects that gave rise to the thought experiment in the first place, and guides us in how to think about their order of importance. The second reason is that even though thought experiments can guide our understanding regardless of whether it is possible to experimentally perform them, it is also important to investigate whether they present practical possibilities. They can act as benchmarks for testing how well one can control the heat and entropy flow in experiments, with the goal being to minimize heat waste in electronics. The Szilard engine, with its measurement and memory scheme, is ideal in this sense.

In the rest of this article we address the two following questions related to how we can limit excitations to the first excited level only, using a simple protocol for the barrier insertion: How sensitive is the nonadiabatic splitting of the wave function to asymmetry in barrier insertion, and what is the probability of exciting states higher than the lowest two when we insert the barrier with a finite rate.

II. ANALYSIS

The box is shown in Fig. 2 and is defined by the potential $V(x) = 0$ for $x \in [-a, b]$ and $V(x) = \infty$ elsewhere. The barrier is a δ function with time-dependent strength $\alpha(t)$ inserted at $x = 0$. We choose the barrier to be a δ function since it allows presenting the eigenstates in analytical form. A barrier with finite width was used in [4], while in [15] they used a δ function barrier and obtained similar results. The width of the barrier would only affect the tunneling rate between the compartments, but the qualitative results would remain unchanged. The insertion of the barrier is described by a time-dependent Hamiltonian given by

$$\hat{H}(t) = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + \alpha(t)\delta(x), \tag{1}$$

where m is the mass of the particle. The instantaneous eigenfunctions $|\psi_n(t)\rangle$ that evolve are found as the solution to the time-independent Schrödinger equation

$$\hat{H}(t)\,|\psi_n(t)\rangle = E(t)\,|\psi_n(t)\rangle\,. \tag{2}$$

At any given time the instantaneous eigenfunctions is an orthonormal set $\langle\psi_n|\psi_m\rangle = \delta_{n,m}$. Therefore the total wave function $|\Psi(t)\rangle$, which is the solution of the time-dependent

Schrödinger equation

$$i\hbar\,\partial_t\,|\Psi(t)\rangle = \hat{H}\,|\Psi(t)\rangle\,, \tag{3}$$

can be expressed as a linear combination of them

$$|\Psi(t)\rangle = \sum_n c_n(t)\,|\psi_n(t)\rangle\,e^{i\theta_n(t)}, \quad \theta_n = -\frac{1}{\hbar}\int_0^t E_n(t')dt'. \tag{4}$$

Here $c_n(t)$ is a set of complex constants satisfying $\sum_n^\infty |c_n(t)|^2 = 1$. As shown in Appendix A, the system of coupled differential equations giving the time evolution of the coefficients $\{c_n\}$ is

$$\dot{c}_n(t) = -\sum_{m\neq n} c_m(t)\frac{\langle\psi_n(t)|\partial_t\hat{H}|\psi_m(t)\rangle}{E_m - E_n}e^{i(\theta_m-\theta_n)}. \tag{5}$$

We first need to find the instantaneous solutions $|\psi_n(t)\rangle$ for the asymmetric barrier problem, and the details of these calculations are given in Appendix B. After finding the instantaneous solutions we numerically solve Eq. (5) to find the time evolution of $|\Psi(t)\rangle$.

III. RESULTS

Let us now see to what extent it is possible to make the probability of finding the particle in either compartment equal (or as close to equal as possible), while limiting excitations to higher energy states.

We set the total length of the box equal to $L = a + b = 1$, and define $a = 1/2 + \epsilon$, where ϵ is the asymmetry parameter that determines how much larger the compartment on the left side of the barrier is than the one on the right side. We also set $\hbar = m = 1$. The initial state is chosen to be the ground state, which is $c_1(0) = 1$ and $c_n(0) = 0$ for $n > 1$. We found that including the six first eigenstates was sufficient to capture all the excitations for the insertion rates we explored. We set the maximum strength of the barrier at the end of the protocol ($t = \tau$) to $\alpha(\tau) = 400\,E_0$, where E_0 is the ground state of the box of $L = 1$ without a barrier. This value was chosen to make sure that the coefficients $\{c_n(\tau)\}$ converged to constant values.

For the protocol we chose $\alpha(t) = At^2$, where A is some constant that determines the rate of insertion. We also tried a linear protocol, but found that in order to limit higher-order excitation the rate of insertion had to start small and steadily increase as a function of time. The reason for this can be understood by studying the coupling between the $\{c_n(t)\}$ in Eq. (5) at a given time t

$$\frac{\langle\psi_n(t)|\partial_t\hat{H}|\psi_m(t)\rangle}{E_m - E_n} = \dot{\alpha}(t)\frac{\langle\psi_n(t)|\hat{\delta}(x)|\psi_m(t)\rangle}{E_m - E_n}. \tag{6}$$

When we insert the barrier, the probability of finding the particle at the insertion point decreases in proportion to the strength of the barrier. Therefore the numerator, $\langle\psi_n(t)|\hat{\delta}(x)|\psi_m(t)\rangle$, which measures overlap between the eigenstates at the insertion point, will be largest in the beginning and decrease toward zero as the barrier strength is increased. This prevents transitions for high barriers. The denominator is the energy difference between the eigenstates, $E_m - E_n$, and its dependence on the barrier strength is shown in Fig. 3. The energy difference between the ground state and the first exited state is largest in the beginning and asymptotically approaches a final small value that increases with the asymmetry between the compartments. This makes transition between these more likely as the barrier strength increases.

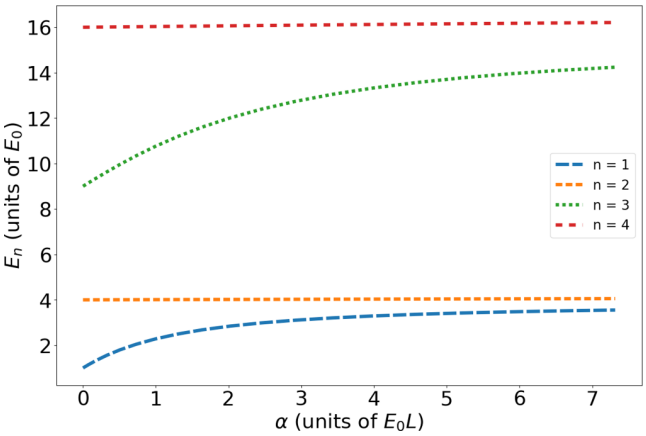

FIG. 3. Energy levels as a function of time. We see that the odd energy levels approach the evens as the strength of the barrier increases, and the final spacing between them decreases with the magnitude of the asymmetry.

In Fig. 4 we plot the ratio $\langle\psi_1(t)|\hat{\delta}(x)|\psi_m(t)\rangle/(E_m - E_1)$ and interpret its magnitude as an indication of the coupling strength between the ground state and the mth eigenstate. As argued in the previous paragraph we see that indeed the ground state's coupling to the first excited state dominates over its coupling to other eigenstates once the barrier has reached a certain strength ($\simeq 4\,E_0$ in this example, where $\epsilon = 0.1$). As seen in Eq. (6), we can control the coupling strength via $\dot{\alpha}(t)$. By choosing a $\dot{\alpha}(t)$ that is small in the beginning and large toward the end of the protocol, we suppress early

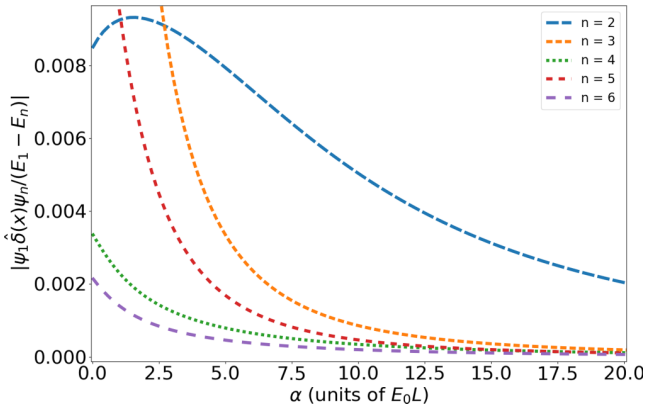

FIG. 4. Dependence of the ratio $\frac{\langle\psi_1(t)|\hat{\delta}(x)|\psi_m(t)\rangle}{E_m - E_1}$ on the strength of the barrier α. Its magnitude gives us an indication of the coupling between the ground state and the higher excited states. We see that the coupling between the ground state and the first excited state remains substantial for high values of α, while all the others decay quickly. This indicates that we can induce transitions between those two levels without exciting higher states when α is large. This plot was obtained with $\epsilon = 0.1$.

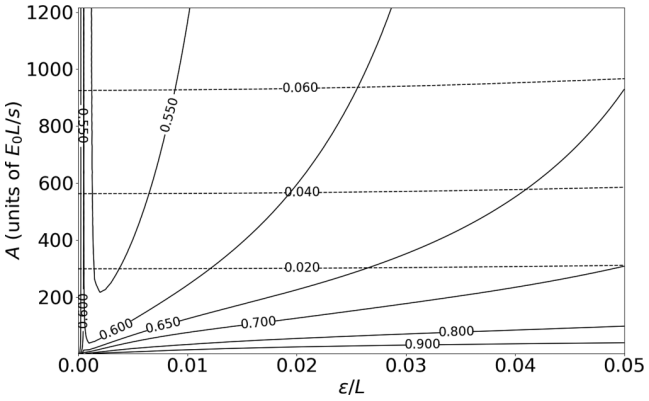

FIG. 5. Probability of finding the particle in the largest compartment (solid lines), as a function of the barrier insertion rate constant A and the asymmetry parameter ϵ. The probability to excite levels higher than the first excited state is shown in the dashed lines.

transitions between the levels when $\langle\psi_n(t)|\hat{\delta}(x)|\psi_m(t)\rangle/(E_m - E_n)$ is large. Since the energy difference between the ground state and the first excited state becomes much smaller than the difference between the ground state and any of the higher states, we can induce transitions between them, even when the wave function overlap is very small, if we choose a $\dot{\alpha}(t)$ that is suitably large.

In Fig. 5 we show a contour plot of the probability of finding the particle in the bigger compartment (solid lines) at the end of the protocol as a function of the asymmetry parameter ϵ/L and the insertion rate parameter A. We see that even for asymmetries of the order of $\epsilon \sim 0.01$ the probability of finding the particle in the bigger compartment is quite large. Although increasing the barrier faster makes the probabilities of finding the particle in either side more equal it also incurs a penalty; the faster you increase the barrier the more likely it is that you excite higher-order states in the energy spectrum. Higher-order excitations increases the entropy of the system, since the internal states of the Szilard engine are assumed to be either the ground state (bigger compartment) or the first excited state (smaller compartment).

IV. SUMMARY AND DISCUSSION

When designing a Szilard engine one wants the probabilities of finding the particle in either compartment after barrier insertion to be equal. Experimentally it might be difficult to design a perfectly symmetric double-well potential. We point out the fact that excitations to the first excited state are special in the sense that after the barrier strength becomes high enough to stop tunneling between the two compartments, and a measurement to determine which compartment the particle is found is performed, the system is still in the ground state for the relevant compartment. This is a generic result, but exactly how to limit the excitations to only the first excited state depends on the specific protocol $\alpha(t)$. We have used a simple protocol that is quadratic in time as an example, and investigated how sensitive the probability distribution of the divided single-particle box is to asymmetry between the compartment size. We find that for this protocol even small differences between the width of the compartments, results in a probability distribution that is skewed toward the larger compartment. The faster one increases the barrier strength, the more even the final distribution becomes. However, this rapid increase also leads to higher-order excitations in the box, which results in unwanted entropy production. The question remains whether a protocol can be constructed such that it gives an equal final distribution between the left and right side, and how sensitive it is to variations in the asymmetry.

ACKNOWLEDGMENT

We thank Y. M. Galperin for valuable discussions and comments on the manuscript.

APPENDIX A: WAVE FUNCTION FOR TIME-DEPENDENT HAMILTONIAN

In this section we follow [16] (Sec. 10.1.2) and write the total wave function $|\Psi(t)\rangle$ as a linear combination of the instantaneous eigenstates $|\psi(t)_n\rangle$ and derive the coupled differential equation for the coefficients. When the Hamiltonian changes with time, the eigenfunctions and eigenvalues are also time dependent,

$$\hat{H}(t)\,|\psi_n(t)\rangle = E_n(t)\,|\psi_n(t)\rangle\,. \tag{A1}$$

The eigenfunctions at any given time is an orthonormal set, $\langle\psi_n(t)|\psi_m(t)\rangle = \delta_{n,m}$, and the total wave function which can be found as the solution of the time-dependent Schrödinger equation

$$i\hbar\,\partial_t\,|\Psi(t)\rangle = \hat{H}\,|\Psi(t)\rangle \tag{A2}$$

can be expressed as a linear combination of them:

$$|\Psi(t)\rangle = \sum_n c_n(t)\,|\psi_n(t)\rangle\,e^{i\theta_n(t)}, \tag{A3}$$

where

$$\theta_n = -\frac{1}{\hbar}\int_0^t E_n(t')dt'. \tag{A4}$$

Inserting this linear combination into the time-dependent Schrödinger equation gives us

$$i\hbar\textstyle\sum_n[\dot{c}_n\,|\psi_n\rangle + c_n\,|\dot{\psi}_n\rangle + ic_n\,|\psi_n\rangle\,\dot{\theta}_n]e^{i\theta_n} \tag{A5}$$

$$= \textstyle\sum_n c_n\hat{H}\,|\psi_n\rangle\,e^{i\theta_n}. \tag{A6}$$

Now since $\dot{\theta}_n = -E_n/\hbar$ and $\hat{H}\,|\psi_n\rangle = E_n\,|\psi_n\rangle$, the right-hand side exactly cancels the last term on the left-hand side and we are left with

$$\sum_n[\dot{c}_n\,|\psi_n\rangle + c_n\,|\dot{\psi}_n\rangle]e^{i\theta_n} = 0. \tag{A7}$$

We now take the inner product with the eigenfunction ψ_m, and since the eigenfunctions constitute an orthonormal set at any given time t, we obtain a set of N coupled differential equations for the N coefficients $c_n,\ n \in [1, N]$.

$$\sum_n[\dot{c}_n\delta_{m,n} + c_n\,\langle\psi_m|\dot{\psi}_n\rangle]e^{i\theta} = 0, \tag{A8}$$

$$\dot{c}_m(t) = -\sum_n c_n\,\langle\psi_m|\dot{\psi}_n\rangle\,e^{i(\theta_n-\theta_m)}. \tag{A9}$$

We can rewrite this equation by taking the time derivative of Eq. (A1) and then the inner product with ψ_m to obtain

$$\langle\psi_m|\dot{\hat{H}}|\psi_n\rangle + E_m\langle\psi_m|\dot{\psi}_n\rangle = \dot{E}\delta_{m,n} + E_n\langle\psi_m|\dot{\psi}_n\rangle, \quad \text{(A10)}$$

which shows us that the inner product $\langle\psi_m|\dot{\psi}_n\rangle$ can be written as

$$\langle\psi_m|\dot{\psi}_n\rangle = \frac{\langle\psi_m|\dot{\hat{H}}|\psi_n\rangle}{E_n - E_m}, \quad \text{(A11)}$$

as long as the system is nondegenerate and $n \neq m$. Putting this result into Eq. (A9) we get

$$\dot{c}_m = -c_m\langle\psi_m|\dot{\psi}_m\rangle - \sum_{n\neq m} c_n \frac{\langle\psi_m|\dot{\hat{H}}|\psi_n\rangle}{E_n - E_m} e^{i(\theta_n-\theta_m)}. \quad \text{(A12)}$$

This form of the differential equation is particularly well suited to our problem. First, the Hamiltonian contains a δ function at $x = 0$, so the integral $\langle\psi_m|\dot{\hat{H}}|\psi_n\rangle$ is simply given by (using the eigenfunctions from Appendix B)

$$\langle\psi_m|\dot{\hat{H}}|\psi_n\rangle = \dot{\alpha}A_nA_m\sin(k_na)\sin(k_ma). \quad \text{(A13)}$$

In addition, the term $\langle\psi_m|\dot{\psi}_m\rangle$ is always zero. This is because the instantaneous eigenfunctions $|\psi_m\rangle$ are orthonormal ($\langle\psi_m|\psi_m\rangle = 1$) and real:

$$\frac{\partial}{\partial t}\langle\psi_m|\psi_m\rangle = \langle\dot{\psi}_m|\psi_m\rangle + \langle\psi_m|\dot{\psi}_m\rangle = 0. \quad \text{(A14)}$$

Since $\langle\psi_m|\dot{\psi}_m\rangle = \langle\dot{\psi}_m|\psi_m\rangle^*$ we get

$$\langle\psi_m|\dot{\psi}_m\rangle = -\langle\psi_m|\dot{\psi}_m\rangle^* \rightarrow \mathrm{Re}[\langle\psi_m|\dot{\psi}_m\rangle] = 0. \quad \text{(A15)}$$

Therefore the coupled differential equations we need to solve become

$$\dot{c}_m = -\sum_{n\neq m} c_n \frac{\langle\psi_m|\dot{\hat{H}}|\psi_n\rangle}{E_n - E_m} e^{i(\theta_n-\theta_m)}. \quad \text{(A16)}$$

APPENDIX B: ASYMMETRIC BARRIER

We can find the stationary states from the time-independent Schrödinger equation and they have the form

$$\psi(x) = \begin{cases} A\sin[k(x+a)], & x \in [-a, 0], \\ B\sin[k(x-b)], & x \in [0, b], \end{cases} \quad \text{(B1)}$$

where $k = \sqrt{2mE}/\hbar$. At $x = 0$ the wave function is continuous while its derivative has a discontinuity. These two conditions are

$$\lim_{\epsilon\to 0}[\psi(0-\epsilon) - \psi(0+\epsilon)] = 0, \quad \text{(B2)}$$

$$\lim_{\epsilon\to 0}[\dot{\psi}(0+\epsilon) - \dot{\psi}(0-\epsilon)] = \frac{2m\alpha}{\hbar^2}\psi(0), \quad \text{(B3)}$$

and for our system they result in

$$A\sin(ka) = -B\sin(kb), \quad \text{(B4)}$$

$$B\cos(kb) - A\cos(ka) = \frac{2m\alpha}{k\hbar^2}A\sin(ka). \quad \text{(B5)}$$

Combining these equations gives us another one, which we can solve numerically to find the wave vectors k for a given a, b, and α.

$$\sin[k(a+b)] = -\frac{2m\alpha}{k\hbar^2}\sin(ka)\sin(kb). \quad \text{(B6)}$$

The solutions to this equation defines a discrete set of allowed values for the wave vector $k \rightarrow k_n$, $n = 1, 2, \ldots$, which determines the energy spectrum of the system via

$$E_n = \frac{\hbar}{2m}k_n^2. \quad \text{(B7)}$$

The wave function has to be normalized on the domain of x,

$$\int_{-a}^{0} A_n^2\sin^2[k_n(x+a)] + \int_0^b B_n^2\sin^2[k_n(x-b)] = 1, \quad \text{(B8)}$$

which combined with Eq. (B4) gives us the normalization constants A_n

$$A_n^2 = \left[\frac{a}{2} - \frac{\sin(2k_na)}{4k_n} + \frac{\sin^2(k_na)}{\sin^2(k_nb)}\left(\frac{b}{2} - \frac{\sin(2k_nb)}{4k_n}\right)\right]^{-1}. \quad \text{(B9)}$$

B_n can be found via Eq. (B4).

[1] L. Szilard, Über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen, Z. Phys. **53**, 840 (1929).

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

[3] S. W. Kim, T. Sagawa, S. De Liberato, and M. Ueda, Quantum Szilard Engine, Phys. Rev. Lett. **106**, 070401 (2011).

[4] J. Gea-Banacloche, Splitting the wave function of a particle in a box, Am J. Phys. **70.3**, 307 (2002).

[5] G. N. Price, S. T. Bannerman, K. Viering, E. Narevicius, and M. G. Raizen, Single-Photon Atomic Cooling, Phys. Rev. Lett. **100**, 093004 (2008).

[6] J. J. Thorn, E. A. Schoene, T. Li, and D. A. Steck, Experimental Realization of an Optical One-Way Barrier for Neutral Atoms, Phys. Rev. Lett. **100**, 240407 (2008).

[7] M. G. Raizen, Comprehensive control of atomic motion, Science **324**, 1403 (2009).

[8] 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).

[9] A. Berut, 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).

[10] V. Serreli, C.-F. Lee, E. R. Kay, and D. A Leigh, A molecular information ratchet, Nature **445**, 523 (2007).

[11] J. V. Koski, V. F. Maisi, J. P. Pekola, and D. V. Averin, Experimental realization of a Szilard engine with a single electron, PNAS **111**, 13786 (2014).

[12] J. V. Koski, A. Kutvonen, I. M. Khaymovich, T. Ala-Nissila, and J. P. Pekola, On-Chip Maxwell's Demon as an Information-Powered Refrigerator, Phys. Rev. Lett. **115**, 260602 (2015).

[13] K. Chida, K. Nishiguchi, G. Yamahata, H. Tanaka, and A. Fujiwara, Thermal-noise suppression in nano-scale Si field-effect transistors by feedback control based on single-electron detection, Appl. Phys. Lett. **107**, 073110 (2015).

[14] M. D. Vidrighin, O. Dahlsten, M. Barbieri, M. S. Kim, V. Vedral, and I. A. Walmsley, Photonic Maxwell's Demon, Phys. Rev. Lett. **116**, 050401 (2016).

[15] S. K. Baek, S. D. Yi, and M. Kim, Particle in a box with a time-dependent δ-function potential, Phys. Rev. A **94**, 052124 (2016).

[16] D. J. Griffiths and D. F. Schroeter, *Introduction to Quantum Mechanics*, 3rd ed. (Cambridge University, New York, 2018).

IV

Deep reinforcement learning for robust quantum optimization

Vegard B. Sørdal[1,*] and Joakim Bergli[1]

[1] *Department of Physics, University of Oslo, 0316 Oslo, Norway*

(Dated: April 10, 2019)

Machine learning techniques based on artificial neural networks have been successfully applied to solve many problems in science. One of the most interesting domains of machine learning, reinforcement learning, has natural applicability for optimization problems in physics. In this work we use deep reinforcement learning and Chopped Random Basis optimization, to solve an optimization problem based on the insertion of an off-center barrier in a quantum Szilard engine. We show that using designed protocols for the time dependence of the barrier strength, we can achieve an equal splitting of the wave function (1/2 probability to find the particle on either side of the barrier) even for an asymmetric Szilard engine in such a way that no information is lost when measuring which side the particle is found. This implies that the asymmetric non-adiabatic Szilard engine can operate with the same efficiency as the traditional Szilard engine, with adiabatic insertion of a central barrier. We compare the two optimization methods, and demonstrate the advantage of reinforcement learning when it comes to constructing robust and noise-resistant protocols.

I. INTRODUCTION

Machine Learning is becoming an essential tool for data analysis and optimization in a wide variety of scientific fields, from molecular [1] and medical science [2] to astronomy [3]. One of the most exiting development in machine learning, comes from combining reinforcement learning [4] with deep neural networks [5]. Reinforcement Learning (RL) differs from supervised and unsupervised learning and is based on letting an agent learn how to behave in a desired way by taking actions in an environment and observing the effect of the action on the environment. In order to define the optimal behavior of the agent, we give it feedback in the form of a reward based on the effect of its previous action. If the action changes the environment into a more desirable state we give it a positive reward, while if it had negative consequences we give it a negative reward. Recently RL has enjoyed increasing popularity in quantum physics, and have been used to explore the quantum speed limit [6, 7], protect qubit systems from noise [8], design new photonic experiments [9], and many other applications [10–12]. For an excellent review of the application of machine learning in physics, see [13].

We use deep reinforcement learning (DRL), specifically Deep-Q Learning (DQL) [5] and Deep Deterministic Policy Gradient (DDPG) [14], to solve an optimization problem based on the barrier insertion of an asymmetric (off-center insertion) quantum Szilard engine, which we will motivate it the following paragraphs. The goal is to find barrier insertion protocols that effectively achieves equal splitting of the wave function of a single-particle-box. We compare the results from DRL with those obtained by using chopped random basis optimization [15], a more traditional optimization algorithm. Finally, since it can be difficult to experimentally determine the exact asymmetry, we show that DRL can be used to find robust protocols, which performs well for a range of asymmetries. We do this by simultaneously training on many instances of the single-particle-box (SPB), where each instance has a different asymmetry. This is essentially the same as training in an environment with a noisy Hamiltonian, as in [7, 8].

The Szilard engine is a classic example of a information processing system, which can convert one bit of Shannon information (obtained by a binary measurement) into an amount $k_BT\ln 2$ of useful work [16]. This is done by inserting a barrier in the center of a SPB, performing a measurement to determine which side of the barrier the particle is found (giving one bit of Shannon information), and then letting the compartment the particle occupies isothermally expand into the empty one resulting in a work-extraction of $k_BT\ln 2$. This work is not free however, since the information obtained has to be stored in a memory, which subsequently has to be deleted at an energy cost of $k_BT\ln 2$ according to Landauer's principle [17]. Both work extraction from a Szilard engine, and Landauer's principle, have recently been experimentally confirmed [18–21].

For the quantum version of the Szilard engine [22], there are some subtle differences in the entropy flow during insertion, expansion, and removal of the barrier [23]. Moreover, the position of the particle is now described by a quantum wave function, which is divided into two parts when inserting the barrier. When adiabatically inserting a barrier in the center of a quantum SPB in its ground state, the wave function is split in half in such a way that each half becomes a new ground state in each compartment, when the barrier strength goes to infinity. The probability to find the particle on either side of the barrier after insertion becomes 1/2. However, as long

* vegardbs@fys.uio.no

as there is an asymmetry in the insertion of the barrier, i.e. it is not put exactly in the center, the adiabatic theorem guarantees that the particle will be found in the larger compartment [24]. Since the initial state is the ground state, and the adiabatic theorem implies the time evolved state will stay in its instantaneous eigenstate, the particle always ends up in the global ground state. The global ground state is found in the larger compartment since the energy is proportional to $L_{R(L)}^{-2}$, where $L_{R(L)}$ is the width of the compartment on the right(left) side of the barrier.

If we want to achieve equal probability on both sides of the barrier for asymmetric insertion, we have to insert the barrier non-adiabatically in such a way that we excite higher eigenstates. This will in general decrease the efficiency of the quantum Szilard engine, since the measurement only determines which side the particle is found, not its exact eigenstate. However, there is one special way of obtaining exact splitting of the wave function without losing any information in the measurement, for the asymmetric Szilard engine [25]: If we insert the barrier in such a way that the total wave function is a superposition of only the first and second eigenstate at the time of measurement, i.e. $|\Psi\rangle = (|\psi_1\rangle + |\psi_2\rangle)/\sqrt{2}$, the which-side measurement does not result in any information loss since the second eigenstate becomes the ground state of the smaller compartment. When one now measure which compartment the particle is in, one is certain that it is in the ground state of the respective compartment.

Our goal is to split the wave function of a single-particle-box in the ground state, by inserting a barrier off-center, in such a way that only the second eigenstate is excited, and the probability to find the particle in all higher states are as close to zero as possible. However, finding a protocol for the barrier insertion which will achieve this goal is non-trivial, since it will have to take advantage of complicated interference between the time-dependent eigenstates.

II. SINGLE-PARTICLE-BOX

The SPB is defined by the potential $V(x) = 0$ for $x \in [-L/2, L/2]$, where L is the total width of the box, and $V(x) = \infty$ elsewhere. The barrier is a δ-function potential inserted at $x = d \geq 0$. An illustration of the SPB is shown in Fig. 1, along with its three first eigenfunctions and eigenenergies before the barrier is inserted. If $\epsilon = 0$ the box is split symmetrically, i.e. the width of the left and right compartment is equal. However, for $d > 0$, the width of the left compartment becomes $L_L = L/2 + d$, while the width of the right compartment becomes $L_R = L/2 - d$. The time-dependent Hamiltonian of the insertion procedure is given by

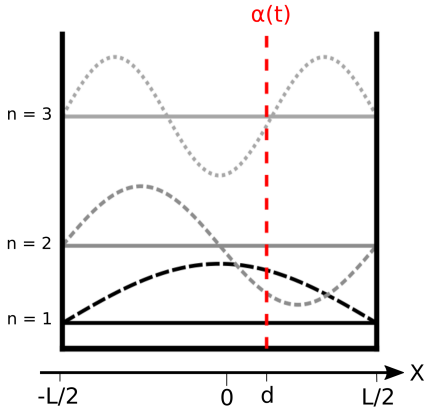

FIG. 1. Illustration of a single particle box with total width L. The eigenfunctions and eigenenergies are shown for the initial state $\alpha(t) = 0$.

$$\hat{H}(t) = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + \alpha(t)\delta(x - d), \tag{1}$$

where $\alpha(t)$ is the strength of the barrier at time t, and m is the mass of the particle. For the rest of this article we set $\hbar = m = 1$. The total wave function, $|\Psi(t)\rangle$, can be expressed as a linear combination of the instantaneous eigenfunctions

$$|\Psi(t)\rangle = \sum_n c_n(t)\,|\psi_n(t)\rangle\, e^{i\theta_n(t)}, \quad \theta_n = -\frac{1}{\hbar}\int_0^t E_n(t')\,dt', \tag{2}$$

where $E_n(t)$ are the instantaneous eigenenergies when the barrier strength is $\alpha(t)$, $|\psi_n(t)\rangle$ are the instantaneous eigenfunctions, and $c_n(t)$ are complex coefficients. The initial state is therefore given by $|c_1(0)|^2 = 1$, and the goal is to construct a protocol $\alpha(t)$, which brings us to a final state where $|c_1(T)|^2 = |c_2(T)|^2 = 1/2$, where T is the duration of the protocol. More details on how the instantaneous eigenstates are calculated, and how the time evolution of the total wave function is numerically solved, is given in [25].

III. CRAB OPTIMIZATION

We use chopped random-basis (CRAB) optimization [15] to find protocols $\alpha(t)$ that splits the wave function in two equal halves for asymmetric barrier insertion in a quantum box. In CRAB optimization we expand the protocol in a complete basis (the Fourier series in our case), in the following way

$$\alpha(t) = \alpha_0(t)\left[1 + \lambda(t)\sum_{n=1}^{N_c} A_n\cos(\omega_n) + B_n\sin(\omega_n)\right]. \tag{3}$$

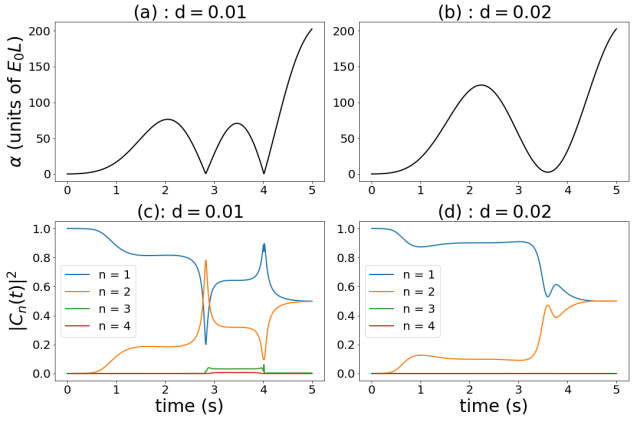

FIG. 2. Results from the CRAB optimization for $d = 0.01$ and $d = 0.02$. In (a) and (b) we show the protocols $\alpha(t)$, while in (c) and (d) we show the time evolution of $|c_n(t)|^2$. We see that the protocol in (b) gives negligible excitations to states $n > 2$ throughout its duration. However the protocol in (a) excites the third eigenstate during the first discontinuity in $\dot{\alpha}(t)$ right before $t = 3$, but this excitation is depleted during the second discontinuity around $t = 4$.

Here $\alpha_0(t)$ is an initial guess for the optimal protocol, $\lambda(t)$ is a regularization function used to implement boundary conditions, and $\{A_n, B_n, \omega_n\}$ is the set of Fourier coefficients we optimize to maximize the cost function

$$C(\{A_n, B_n, \omega_n\}) = 1 - \sum_{n=1}^{2} \left(|c_n(T)|^2 - 0.5\right)^2. \quad (4)$$

We fix the boundary conditions to be $\alpha(0) = 0$ and $\alpha(T) = 200E_0L$ (where $E_0 = \pi^2/2$ is the ground state at $\alpha(0) = 0$) , and choose $\lambda(t) = \sin(\pi t/T)$. To minimize Eq. (4) we use a gradient free method, like the Nelder-Mead [26] or Powell's method [27]. Using the Nelder-Mead method we are able to almost exactly split the wave function in half, and results for $d = 0.01$ and $d = 0.02$ are shown in Fig. 2. In these examples we obtained $|c_1(T)|^2 = 0.4986$, $|c_2(T)|^2 = 0.4979$, and $\sum_{n>2} |c_n(T)| \simeq 10^{-3}$ for $d = 0.01$, while for $d = 0.02$ we got $|c_1(T)|^2 = 0.5001$, $|c_2(T)|^2 = 0.4999$, and $\sum_{n>2} |c_n(T)|^2 \simeq 10^{-5}$. In Fig. 2(a) and Fig. 2(b) we show example protocols for $d = 0.01$ and $d = 0.02$, respectively, while in Fig. 2(c) and Fig. 2(d) we show the time evolution of the probability to be in a given eigenstate, $|c_n(t)|^2$.

The protocols obtained by CRAB are designed to split the wave function in two for a given asymmetry. They work extremely well for the asymmetry they were designed for. However, the protocols generalize poorly to other asymmetries, as shown in Fig. 3. There we plot $|c_n(T)|^2$ as a function of the asymmetry d, using the protocol designed for $d = 0.01$ and $d = 0.02$. We see that the performance of a protocol designed for a specific asymmetry dramatically reduces if it is applied to single-particle-boxes of different asymmetries. An interesting feature is seen in Fig. 3(b), where the protocol designed for $d = 0.02$ achieves exact splitting for asymmetries other than the one that was used for training. However, even this protocol has bad performance in the regions between these points of exact splitting, so it would not be useful unless one knows the exact asymmetry of the single-particle-box.

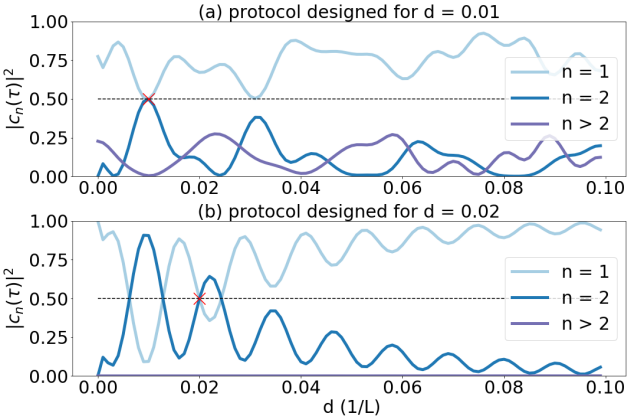

FIG. 3. Plot showing how the protocols designed for two specific asymmetries performs on other asymmetries. In (a) we show the results for the protocol designed for $d = 0.01$, while in (b) we show the one designed for $d = 0.02$. The light blue and the blue line shows the occupation at $t = T$ for the first and second eigenstate respectively, while the purple line shows the occupation of all eigenstates higher than the second, i.e. the unwanted excitations. The black dashed lines shows the target $|c_n(T)|^2 = 0.5$, and the red crosses shows the asymmetry trained on.

IV. DEEP Q-LEARNING

We now give a short review of the DQL algorithm introduced in [5]. In the next section we will show how this general algorithm is adapted to our problem. A schematic of the basic reinforcement learning protocol is shown in Fig. 4. At time t the environment is in a given state s_t. The agent performs an action a_t which induces a state change of the environment from s_t to s_{t+1}. The agent then receives an observation of the new state of the environment, s_{t+1}. After taking an action the agent receives a reward $r_t = r(s_t, a_t, s_{t+1})$. The reward function $r(s_t, a_t, s_{t+1})$ is designed by us, according to what goal we want the agent to achieve.

The behavior of the agent is determined by its policy $\pi(a_t|s_t)$, which is the probability of taking the action a_t in given the observation s_t. If the agent is in state s_t, the Q-function (quality function) $Q_\pi(s_t, a_t)$ gives the expected cumulative reward given that the action a_t is

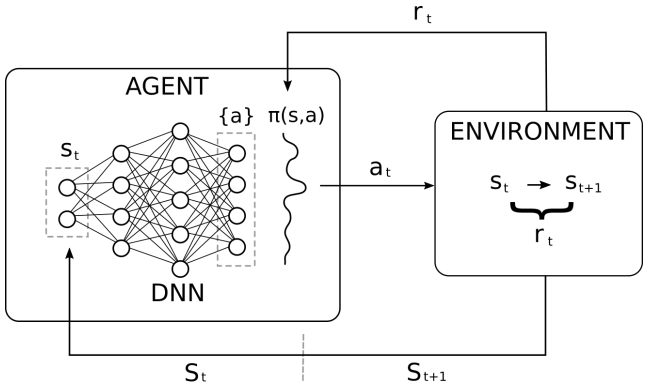

FIG. 4. Schematic showing the basic setup of deep Q-learning. The current state of the system s_t is fed as input nodes into a deep neural network (DNN). The output nodes are the set of all possible actions, $\{a\}$, and their values are the estimated Q-value for the given state-action pair. The policy $\pi(s,a)$ is given by the action node with the highest output value, or by a random action if the agent is exploring. The action determined by the policy is performed in the environment, inducing a state change from $s_t \to s_{t+1}$. Associated with this state change, a reward r_t is given, which is used to determine how good the given action was in this state. This reward is fed back into the DNN and used to update its weights according to Eq. (8). Schematic adapted from [28].

performed and the policy π is followed for all proceeding states.

$$\begin{aligned} Q(s_t,a_t) &= \mathrm{E}_{s_{t+1}}\left[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \ldots | s_t, a_t, \pi\right] \\ &= \mathrm{E}_{s_{t+1}}\left[r_t + \gamma Q(s_{t+1}, a_{t+1}) | s_t, a_t, \pi\right] \end{aligned} \tag{5}$$

Here $\gamma \leq 1$ is a discount parameter, which determines how much the agent values immediate reward compared to future reward. If $\gamma < 1$ the agent will value future reward less than immediate reward, which is useful for learning in stochastic environments where the future is more uncertain. The optimal Q-function, $Q^*_\pi(s_t,a_t)$, is the maximum expected cumulative reward obtained by taking the action a_t in state s_t and then acting optimally thereafter, and it is shown to obey the Bellman optimality equation [4]

$$Q^*_\pi(s_t,a_t) = \mathrm{E}_{s_{t+1}}\left[r_t + \gamma \max_{a_{t+1}} Q^*_\pi(s_{t+1},a_{t+1}) | s_t, a_t\right] \tag{6}$$

If we have $Q^*_\pi(s,a)$ for all possible state-action pairs, it is clear that we can find the optimal policy, π^*, by choosing $a_t = \arg\max_{a'} Q^*_\pi(s_t,a')$, i.e. following the policy

$$\pi^*(a_t|s_t) = \arg\max_{a'} Q^*_\pi(s_t,a'). \tag{7}$$

The key idea introduced in [5], is to estimate the optimal Q-function using a neural network $Q^*_\pi(s,a) \simeq Q^*_\pi(s,a,\theta)$, where θ is the weights and biases of the neural network. This neural network is called a Deep-Q network (DQN), and is updated by performing gradient ascent on the mean-squared-error of the current predicted $Q^*_\pi(s,a,\theta)$, while using the Bellman equation as the target. The loss function for DQN is therefore

$$L(\theta) = E_{s_{t+1}}\left[\left(Q^*_\pi(s_t,a_t,\theta) - y_t\right)^2\right]. \tag{8}$$

where

$$y_t = r_t + \gamma \max_{a_{t+1}} Q^*_\pi(s_{t+1},a_{t+1},\theta) \tag{9}$$

To create the neural network we used tensorflow's implementation of the Keras API [29, 30], with Adam [31] as the optimizer. The network consists of three hidden layers, with 24, 48 and 24 neurons, respectively, as well as 2 input neurons and 20 output neurons. When the network is initialized its predictions for the optimal Q^*_π-values are of course totally wrong. So if we always chose the actions that maximizes the current predicted Q^*_π-values, the agent would not learn anything. We need to let the agent explore the state-action space by randomly performing actions. A typical exploration policy is the ϵ-greedy policy. The agent chooses random actions with probability ϵ, or the ones with the highest Q^*_π-value (greedily) with probability $1-\epsilon$. As time goes and the agent explores more of the environment, ϵ is decreased so that it focuses more on the areas of the state-action space with higher Q^*_π-values by taking deterministic actions. Typically we start by taking completely random actions, $\epsilon = 1$, and let ϵ converge to some finite number $\epsilon \sim 0.05$, so that there is always some exploration going on. As seen in Eq. (8) a single update of the network weights requires the following input: the current state s_t, the action chosen a_t, the immediate reward r_t, and the next state s_{t+1}. We call this tuple, $e_t = (s_t, a_t, r_t, s_{t+1})$, that the network trains on an experience. Instead of training on consecutive experiences we store them all in a memory $M_N = \{e_0, e_1, \ldots, e_N\}$, and then train on randomly drawn batches of samples from the memory. The memory have a finite capacity, and new experiences replace older ones when the memory is full. There are three main advantages of training on the memory: It is data efficient since a single experience can be drawn many times. Only training on consecutive experiences is inefficient, since the network tends to forget previous experiences by overwriting them with new experiences. The time-correlation of consecutive experiences means that the network update due to the current experience determines what the next experience will be, so training can be dominated by experiences from a certain area in the state-action space. Finally we see that in Eq. (8) the current weights of the network determines both the target Q^*_π-value and the predicted Q^*_π-value from the Bellman equation. Thus every network update changes the target Q^*_π-value that we are trying to reach, and makes it hard for the network weights to converge. A simple way to circumvent this problem is to use two neural networks,

one for the target Q^*_π-value (θ^-), and one for the current Q^*_π-value (θ). The target network is softly updated during training according to $\theta^- \leftarrow \theta^-(1-\tau) + \theta\tau$, where τ is a hyper-parameter that determines how close the two networks are in the network parameter space.

V. DQL RESULTS

For our system, we defined the state to be a tuple of the strength of the δ-barrier and the time t, i.e. $S = \{\alpha(t), t\}$. The available actions is a set of $\dot{\alpha}(t)$, given by

$$A = \{\, \dot{\alpha}^{\pm}_n(t) = \pm\, 2^n, \quad \text{for} \quad n = 1, 2, \ldots, 10 \,\}\,.$$

The initial state is $S = \{\alpha(t) = 0, t = 0\}$, and the goal is to reach a state where $|c_0|^2 = |c_1|^2 = 1/2$, at the end of the protocol $t = T$. A sequence of selected actions, from time $t = 0$ to $t = T$, defines a protocol $\alpha(t)$. The number of times the agent chooses an action per protocol is given by N_t, and the time-step is therefore $dt = T/N_t$. The environment that the agent acts in is the quantum mechanical SPB, with initial state $|c_1|^2 = 1$ and $|c_{n>1}|^2 = 0$, and time evolution given by the Schrödinger equation $i\partial_t |\psi(t)\rangle = H(t)\,|\psi(t)\rangle$, which we solve as in [25]. The sequential process for one episode is then

1. Initial state is $s_0 : (\alpha_0 = 0, t = 0)$
2. Agent chooses action based on s_0, e.g. $a_0 = \dot{\alpha}^+_3$
3. The next state is then $s_1 : (\alpha_0 + \dot{\alpha}^+_3 dt, t + dt)$
4. Repeat 2. $\rightarrow$ 3., for $s_1, s_2, \ldots$ until $t = T$.
5. Solve the Schrödinger equation for the given protocol (set of all states $\{\, s_n, t_n \,\}$) and calculate reward. Repeat from 1. until maximum number of episodes reached.

The reward function we used is defined by

$$r(t) = \begin{cases} 0, & \text{if } t < T \text{ and } \alpha \in [0, \alpha_{max}] \\ -10, & \text{if } t < T \text{ and } \alpha \notin [0, \alpha_{max}] \\ 100 \exp\left(- \sum_{n=1}^{2} \frac{\left(|c_n(T)|^2 - 0.5\right)^2}{\sigma} \right), & \text{if } t = T \end{cases} \tag{10}$$

where σ determines how sharp we want the reward distribution to be. If the agent chooses actions such that $\alpha(t) < 0$, we give it a punishment of -10 and set $\alpha(t) = 0$, and for actions that would give $\alpha(t) > \alpha_{max}$ we punish and set $\alpha(t) = \alpha_{max}$. We do this to keep the state space bounded. The space of possible protocols grows exponentially with dt^{-1}, so it is impractical to set dt so small that we get approximately continuous $\dot{\alpha}(t)$. The accuracy of our numerical solution of the quantum time evolution decreases if we have discontinuous $\dot{\alpha}(t)$, so to circumvent this problem we use cubic spline to interpolate the final protocol before calculating the reward.

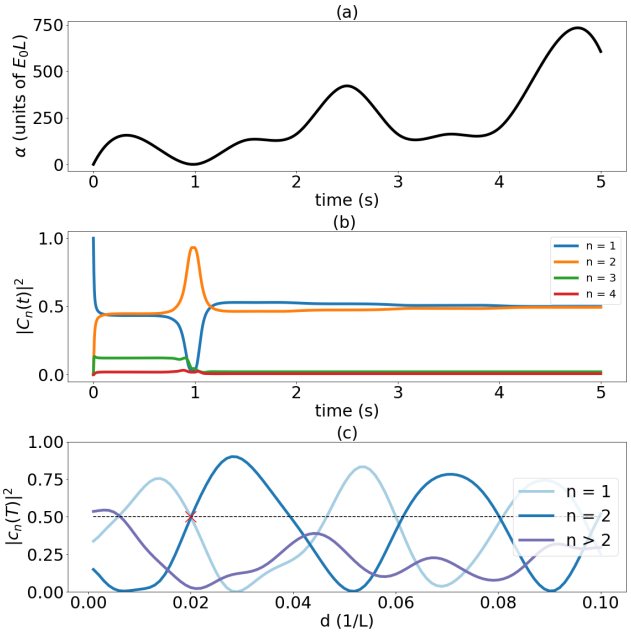

FIG. 5. Results from DQL when training on a single asymmetry, $\epsilon = 0.02$. In (a) we show the protocol $\alpha(t)$, while in (b) we show the time evolution of $|c_n(t)|^2$ for the asymmetry we trained on. Similarly to the protocol in Fig. 2(a), there is a good amount of excitations to the third eigenstate in the very beginning of the protocol, which is then depleted around $t = 1$ s. In (c) we show how the protocol generalizes to other asymmetries, by plotting the distribution $|c_n(T)|^2$ at $t = T$ for asymmetries in the range $d \in [0.01, 0.1]$. The parameters of this protocol was $T = 5$ s, $N_t = 10$, $\alpha_{max} = 800$ $E_0 L$, and $\sigma = 0.05$.

In Fig. 5 we show an example protocol learned by the DQL agent, and the corresponding time evolution of $|c_n(t)|^2$, when training on a single asymmetry ($\epsilon = 0.02$) for 10 000 episodes. The final distribution was $|c_1(T)|^2 = 0.4996$, $|c_2(T)|^2 = 0.4935$, and with higher excitations $\sum_{n>2} |c_n|^2 \simeq 10^{-2}$. The results, when training on a single asymmetry, tended to be worse for DQL than for direct CRAB optimization. There are many ways to improve the results obtained by DQL; we can add actions to, or change the action space, train for a longer time or increase the number of actions per episode N_t. Alternatively, one could implement algorithms similar to DQL that can perform actions in a continuous action space, like deep deterministic policy gradient (DDPG) [14]. However, most of these changes would also increase the necessary training time.

In real experiments one may not know exactly how large the asymmetry of the single-particle-box is. A far more useful protocol would be a robust one, designed to work best for a given range of asymmetries. One of the main benefits of DQL is that it is a model-free algorithm, so this task is easily achieved. One only has to let the agent train on random samples of the set of asymmetries

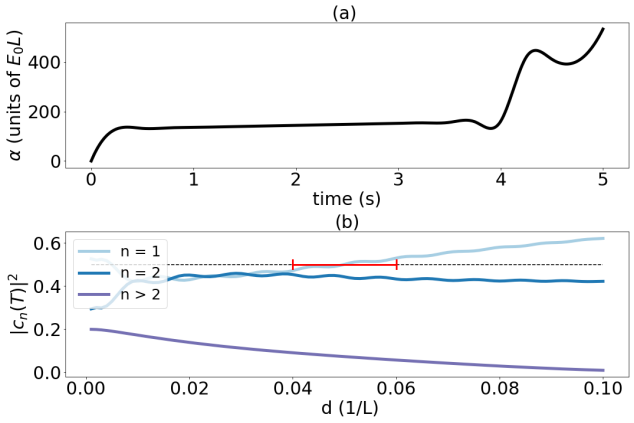

FIG. 6. Results from DQL when training on 10 different asymmetries in the range $d \in [0.04, 0.06]$. In (a) we show the protocol obtained, while in (b) we show $|c_n(T)|^2$ all asymmetries up to $d = 0.1$, where the red bar indicates the range of asymmetries we trained on. When compared to Fig. 3, we see that the protocol performs much better overall than the ones designed for one specific asymmetry, particularly in the range we trained on. The parameters of this protocol was $T = 5\ s$, $N_t = 20$, $\alpha_{max} = 800\ E_0L$, and $\sigma = 0.05$.

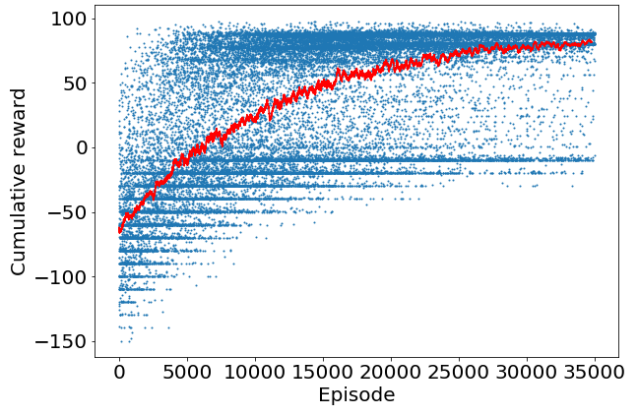

FIG. 7. Scatter plot of the reward received per episode, when training on multiple asymmetries, shown in blue dots, and a running average shown in red. The probability to take random actions is gradually reduced with the number of episodes, leading to a final protocol which the agent determines to be the best.

one wants the protocol to be optimized to. Since the agent tries to maximize the expected cumulative reward, this added stochasticity is no hindrance. How much the agent values a given state-action pair is averaged over the random samples from the memory, which is proportionally filled with the number of asymmetries we train on.

As an example, say one could determine the asymmetry with a given accuracy $d = 0.05 \pm 0.01$. An example protocol that was obtained when training on multiple asymmetries (10 equally spaced samples in the range $d \in [0.04, 0.06]$) is shown in Fig. 6. As seen in Fig. 6(b), this protocol performs better on the full range of asymmetries than the ones designed for a single asymmetry, shown in Fig. 3. The excitation to states higher than the two first eigenstates is largest for small asymmetries. This is due to the fact that when $d \to 0$, the wall is inserted close to the central node of the second eigenstate, and the central anti-node of the third eigenstate, as shown in Fig. 1. Therefore excitations to the second eigenstate becomes less likely, while the opposite is true for excitations to the third eigenstate. Since this is an intrinsic property of the system, it is impossible to find protocols that avoids excitations for $d \to 0$. For $d = 0$, the ground state of the left and right compartment constitute a doubly degenerate global ground state, and to achieve an equal splitting of the wave function, one has to insert the barrier adiabatically [25].

In Fig. (7) we see the total cumulative reward received per episode in a scatter plot, as well as a running average. We see that in the early episodes, where there agent mostly performs random actions, there are many episodes with negative cumulative reward. This is because there is an equal probability that the agent chooses negative and positive $\dot{\alpha}$, and since the initial state is $\alpha(t = 0) = 0$ there is a high probability that the agent chooses actions which gives $\alpha(t) < 0$, resulting in a punishment of -10 every time. In this early stage the agent explores and learns about its environment. As the probability to take random actions decreases (according to the ϵ-greedy protocol) with each episode, the agent takes more deterministic actions based on its experience, and the reward per episode increases steadily. The stochasticity observed in the rewards for final episodes is due a finite final exploration rate $\epsilon = 0.05$. We obtain the final protocol after training by setting $\epsilon = 0$, and let the agent act deterministically. The efficiency of the protocol obtained by training on a range of asymmetries can be increased by implementing the same changes as for the one designed for a single asymmetry.

VI. DEEP DETERMINISTIC POLICY GRADIENT

Our set of possible actions for the DQL algorithm is somewhat arbitrarily chosen. For our specific control problem, there are infinitely many protocols that can achieve our goal, so the exact set chosen is not critically important. However, the performance of the algorithm depends on this choice, and the optimal protocols we find can always be defined by some subset of the total action-space. That is, not all actions are used for the optimal protocol, so we could retroactively reduce the action-space after learning which actions was needed. For many control problems in physics, it is more natural

to let the action values be drawn from a continuous set, on some interval $A \in [a_{min}, a_{max}]$. For DQL, this is not possible, since the optimal policy $\pi^*(a_t|s_t)$ comes from taking the maximum argument of a finite dimensional $Q^*(s_t, a_t)$.

When the action-space is continuous, the optimal Q-function $Q^*(s, a)$ is assumed to be differentiable with respect to the action a. In Deep Deterministic Policy Gradient [14], the goal is to find a deterministic policy $\mu(s)$, which gives is the optimal action to take for any state, $a^* = \mu(s)$. This deterministic policy is approximated by another neural network $\mu(s) \simeq \mu(s, \phi)$, where ϕ are the parameters of the network. The Q-function is, as in DQL, also approximated by a neural network, and the essence of introducing the deterministic policy is to replace the largest Q-value for a state-action pair in the following way:

$$\arg\max_{a'} \; Q^*_\pi(s_t, a', \theta) \rightarrow Q^*(s_{t+1}, \mu(s_{t+1}, \phi), \theta). \quad (11)$$

The Q-network is updated in the same way as in DQL, by using the Bellman equation, but instead of Eq. (9), the target for the loss function now becomes

$$y_t = r_t + \gamma Q^*(s_{t+1}, \mu(s_{t+1}, \phi), \theta). \quad (12)$$

As for the policy network, it was shown in [32] that its weights can updated in proportion to the gradient of the Q-function

$$\phi_{k+1} = \phi_k + \lambda E_{s \in B} \left[\nabla_\phi Q^*(s, \mu(s, \phi), \theta) \right], \quad (13)$$

where λ is the learning rate, which determines the step-size of the gradient ascent. Since the gradient will, in general, move the weights in different directions for different states, an average over a batch of experiences is taken. By applying the chain rule to Eq. 13, we can decompose it into a product of the gradient of the policy with respect to its network weight, and the gradient of the Q-function with respect to the actions

$$\nabla_\phi Q^*(s, \mu(s, \phi), \theta) = \nabla_\phi \mu(s, \phi) \nabla_a Q^*(s, a, \phi)|_{a=\mu(s,\phi)}. \quad (14)$$

Exploration in DDPG is driven by adding noise to the policy, sampled form some distribution N_t suited to the environment, which is annealed over time

$$\mu'(s_t) = \mu(s_t, \phi) + N_t. \quad (15)$$

We use a Gaussian white noise process, and annealed its standard deviation from $\sigma_N = 0.3$ to $\sigma_N = 10^{-4}$ over the course of the training. DPPG is called an actor-critic model, and the sense is that the policy is an actor, taking actions in an environment, and the Q-function acts as a critic, determining how good the actions where, and feeding the result back to the actor.

For the DPPG algorithm, we used an adapted implementation from Keras-RL [33], which includes the same modifications we used for DQL; i.e. experience replay and different networks for the target and current Q-function and policy. The policy network takes as input the same state tuple as for DQL, $S = \{\alpha(t), t\}$, which is connected to three hidden layers, with the same architecture as DQL; 24, 48, and 24 neurons, respectively, and outputs a single value, the action $\dot{\alpha}(t)$. The Q-function network takes as input the action value suggested by the policy, as well as the state $S = \{\alpha(t), t\}$, again connected to three hidden layers with the same architecture as DQL, and outputs a value which is its estimation of the optimal Q-value of the state-action pair. The output actions form the policy network are clipped at $|\dot{\alpha}(t)| \leq 1000\ E_0 L/s$, and we use the same reward function as for DQL.

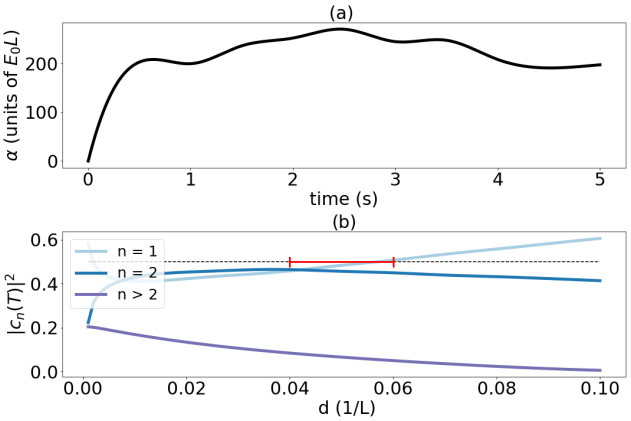

FIG. 8. Results from DDPG, when training on 10 different asymmetries in the range $d \in [0.04, 0.06]$. In (a) we show the protocol itself, while in (b) we show how the protocol performs on a range of asymmetries $d \in [0, 0.1]$. The red bar marks the range we trained on. The parameters of this protocol was $T = 5\ s$, $N_t = 20$, $\alpha_{max} = 800\ E_0 L$, and $\sigma = 0.05$, and we trained for 20 000 episodes.

In Fig. 8(a) we show a protocol obtained from DDPG, when training on 10 asymmetries in the range $d \in [0.04, 0.06]$, and in Fig. 8(b) the performance of the protocol on a range of asymmetries from $d \in [0, 0.1]$. As expected, the best results are obtained for the range of asymmetries we trained on, indicated by a red bar. A rigorous comparison between DQL and DPPG is difficult, partly due to the large amount of hyper-parameter tweaking needed to optimize each algorithm, but largely due to the arbitrary choice of discrete action values for DQL: for our example problem, there is no natural set of available actions to choose. As mentioned earlier, the performance of DQL for our problem, depends on the set of actions chosen, and therefore a fair comparison of the algorithms is complicated. The choice between discrete and continuous-action algorithms, has to be taken based on the specific problem one wants to solve. For our SPB problem, there are infinitely many "good" solutions,

and since we interpolate the protocol at the end of each episode, both DQL and DPPG are well suited.

We used a 3.40 GHz CPU, and the training time for the most resource-intensive computation (the protocol in Fig. (6)) was about 48 hours, so increased training time is something that more advanced computation systems can handle. The most computationally-intensive part of the training, by a large margin, was solving the Schrödinger equation after each episode. As for the hyper-parameters of the neural networks, we used a learning rate $\lambda = 10^{-3}$, target network update every $\tau = 10^{-3}$ time-step, and a replay memory size between $10\% - 50\%$ of the total number of experiences. The ε-greedy exploration policy was a linear decrease from $\varepsilon = 1$ to $\varepsilon = 0.05$.

VII. DISCUSSION AND SUMMARY

We have used CRAB optimization and deep reinforcement learning to construct protocols, $\alpha(t)$, for the time-dependent strength of a barrier inserted asymmetrically in a single-particle-box, in such a way that the wave function is split in two equal halves. These results implies that the asymmetric quantum Szilard engine can reach the same efficiency in information-to-work conversion as the symmetric one, since no information is lost in the which-side measurement.

Using CRAB optimization, the protocols we obtain performs very well for the specific asymmetry we optimize for, but the protocol generalize poorly for different asymmetries. Although more time consuming and than CRAB optimization, we can also use DRL to find high performing protocols when training on single asymmetries. However, one of the biggest strengths of reinforcement learning based techniques is the possibility to perform robust and noise-resistant optimization. When training on a range of different asymmetries simultaneously, DRL can be used to find the protocols that performs best on the average of all the asymmetries sampled. Both DQL and DDPG were able to find good protocols for our example SPB problem, but in general, the choice between discrete and continuous-action algorithms has to be made on the basis of what specific problem one wants to solve. The advantage of using reinforcement learning for quantum control, is multi-faceted: having model-free algorithms makes it simple to change the optimization criterion to make the agent solve different problems within the same environment, one only have to change the reward function to suit the new goal. Furthermore, since the agent is not tailored to any specific environment, it can easily be adopted to work in entirely different systems (e.g. we can use the agents constructed here to perform state-transfer in qubit systems [6]). Finally, the stochastic nature of the agents learning procedure is advantageous when one wants to perform robust optimization which can perform well with noise. These points all suggests that reinforcement learning can become a very useful tool in physics.

ACKNOWLEDGEMENTS

We thank Y. M. Galperin for reading the manuscript and making valuable comments.

[1] Keith T Butler, Daniel W Davies, Hugh Cartwright, Olexandr Isayev, and Aron Walsh, "Machine learning for molecular and materials science," Nature **559**, 547 (2018).

[2] Dinggang Shen, Guorong Wu, and Heung-Il Suk, "Deep learning in medical image analysis," Annual review of biomedical engineering **19**, 221–248 (2017).

[3] JR Primack, A Dekel, DC Koo, S Lapiner, D Ceverino, RC Simons, GF Snyder, M Bernardi, Z Chen, H Domínguez-Sánchez, *et al.*, "Deep learning identifies high-z galaxies in a central blue nugget phase in a characteristic mass range," The Astrophysical Journal **858**, 114 (2018).

[4] Richard Bellman, "A markovian decision process," Journal of Mathematics and Mechanics **6**, 679–684 (1957).

[5] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis, "Human-level control through deep reinforcement learning," Nature **518**, 529–533 (2015).

[6] Marin Bukov, Alexandre GR Day, Dries Sels, Phillip Weinberg, Anatoli Polkovnikov, and Pankaj Mehta, "Reinforcement learning in different phases of quantum control," Physical Review X **8**, 031086 (2018).

[7] Xiao-Ming Zhang, Zi-Wei Cui, Xin Wang, and Man-Hong Yung, "Automatic spin-chain learning to explore the quantum speed limit," Physical Review A **97**, 052333 (2018).

[8] Thomas Fösel, Petru Tighineanu, Talitha Weiss, and Florian Marquardt, "Reinforcement learning with neural networks for quantum feedback," Physical Review X **8**, 031084 (2018).

[9] Alexey A Melnikov, Hendrik Poulsen Nautrup, Mario Krenn, Vedran Dunjko, Markus Tiersch, Anton Zeilinger, and Hans J Briegel, "Active learning machine learns to create new quantum experiments," Proceedings of the National Academy of Sciences **115**, 1221–1226 (2018).

[10] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd, "Quantum machine learning," Nature **549**, 195 (2017).

[11] Daoyi Dong, Chunlin Chen, Hanxiong Li, and Tzyh-Jong Tarn, "Quantum reinforcement learning," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **38**, 1207–1220 (2008).

[12] Lucas Lamata, "Basic protocols in quantum reinforcement learning with superconducting circuits," Scientific reports **7**, 1609 (2017).

[13] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre GR Day, Clint Richardson, Charles K Fisher, and David J Schwab, "A high-bias, low-variance introduction to machine learning for physicists," arXiv preprint arXiv:1803.08823 (2018).

[14] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra, "Continuous control with deep reinforcement learning," arXiv preprint arXiv:1509.02971 (2015).

[15] Tommaso Caneva, Tommaso Calarco, and Simone Montangero, "Chopped random-basis quantum optimization," Phys. Rev. A **84**, 022326 (2011).

[16] L. Szilard, "über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen," Zeitschrift für Physik **53**, 840–856 (1929).

[17] Rolf Landauer, "Irreversibility and heat generation in the computing process," IBM journal of research and development **5**, 183–191 (1961).

[18] Shoichi Toyabe, Takahiro Sagawa, Masahito Ueda, Eiro Muneyuki, and Masaki Sano, "Experimental demonstration of information-to-energy conversion and validation of the generalized jarzynski equality," Nature physics **6**, 988 (2010).

[19] Antoine Bérut, Artak Arakelyan, Artyom Petrosyan, Sergio Ciliberto, Raoul Dillenschneider, and Eric Lutz, "Experimental verification of landauers principle linking information and thermodynamics," Nature **483**, 187 (2012).

[20] Jonne V Koski, Ville F Maisi, Jukka P Pekola, and Dmitri V Averin, "Experimental realization of a szilard engine with a single electron," Proceedings of the National Academy of Sciences **111**, 13786–13789 (2014).

[21] Mihai D Vidrighin, Oscar Dahlsten, Marco Barbieri, MS Kim, Vlatko Vedral, and Ian A Walmsley, "Photonic maxwells demon," Physical review letters **116**, 050401 (2016).

[22] Seth Lloyd, "Quantum-mechanical maxwells demon," Physical Review A **56**, 3374 (1997).

[23] Sang Wook Kim, Takahiro Sagawa, Simone De Liberato, and Masahito Ueda, "Quantum szilard engine," Physical review letters **106**, 070401 (2011).

[24] Julio Gea-Banacloche, "Splitting the wave function of a particle in a box." Americal Journal of Physics **70.3**, 307–312 (2002).

[25] Vegard B. Sørdal and Joakim Bergli, "Quantum particle in a split box: Excitations to the ground state," Phys. Rev. A **99**, 022121 (2019).

[26] John A Nelder and Roger Mead, "A simplex method for function minimization," The computer journal **7**, 308–313 (1965).

[27] Michael JD Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," The computer journal **7**, 155–162 (1964).

[28] Hongzi Mao, Mohammad Alizadeh, Ishai Menache, and Srikanth Kandula, "Resource management with deep reinforcement learning," in *Proceedings of the 15th ACM Workshop on Hot Topics in Networks* (ACM, 2016) pp. 50–56.

[29] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," (2015), software available from tensorflow.org.

[30] François Chollet *et al.*, "Keras," `https://keras.io` (2015).

[31] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).

[32] David Silver, Guy Lever, Nicolas Heess, Thomas Degris, Daan Wierstra, and Martin Riedmiller, "Deterministic policy gradient algorithms," in *ICML* (2014).

[33] Matthias Plappert, "keras-rl," `https://github.com/keras-rl/keras-rl` (2016).