DTU Library

A Friendly Guide to Exorcising Maxwell's Demon

de Oliveira Junior, A.; Brask, Jonatan Bohr; Chaves, Rafael

Published in:
PRX Quantum

Link to article, DOI:
10.1103/phkv-wrsd

Publication date:
2025

Document Version
Publisher's PDF, also known as Version of record

Tutorial

A Friendly Guide to Exorcising Maxwell's Demon

A. de Oliveira Junior,[1,*] Jonatan Bohr Brask,[1] and Rafael Chaves[2,3]

[1] *Center for Macroscopic Quantum States bigQ, Department of Physics, Technical University of Denmark, Fysikvej 307, 2800 Kgs. Lyngby, Denmark*

[2] *International Institute of Physics, Federal University of Rio Grande do Norte, 59078-970 Natal, RN, Brazil*

[3] *School of Science and Technology, Federal University of Rio Grande do Norte, Natal, Brazil*

(Received 16 March 2025; revised 2 June 2025; published 11 August 2025)

The birth, life, and death of Maxwell's demon provoked a profound discussion about the interplay between thermodynamics, computation, and information. Even after its resolution, the demon continues to inspire a multidisciplinary field. This tutorial offers a comprehensive overview of Maxwell's demon and its enduring influence, bridging classical concepts with modern insights in thermodynamics, information theory, and quantum mechanics.

DOI: 10.1103/phkv-wrsd

CONTENTS

*Contact author: alexssandredeoliveira@gmail.com

I. INTRODUCTION

Thought experiments in physics are the birthplace of new paradigms. Einstein's relativity defied Newtonian notions of space and time [1,2] and led him to have what he later called his "happiest thought," when he conceived the equivalence principle [3,4]. Meanwhile, Schrödinger's cat [5] illustrates the strangeness of quantum mechanics. Many of these *mental exercises* have evolved from theoretical explorations to experimental realizations. When Einstein questioned whether quantum mechanics offered a complete description of reality [6], John Bell provided an elegant answer decades later with his groundbreaking theorem [7]. Not long after, experiments by Alain Aspect, Jean Dalibard, and Gérard Roger confirmed Bell's predictions [8]—bringing such thought experiments to life as tangible realities. Yet, some thought experiments remain unresolved, such as Wigner's friend [9], which highlights the need for a deeper interpretation of quantum mechanics (see Ref. [10] for a state-of-the-art overview related to the topic).

This tutorial discusses a thought experiment that took more than a century of scientific development to be resolved. By the 19th century, it was commonly understood that thermodynamic fluctuations at the particle level could lead to apparent violations of the second law of thermodynamics. To illustrate how these fluctuations could be systematically exploited to seemingly break the second law, Maxwell proposed a thought experiment, now famously known as Maxwell's demon [11,12].

Maxwell envisioned a scenario in which, by knowing the positions and velocities of gas particles, one could reduce the entropy of the system without performing any work, thus appearing to defy the second law. More specifically, the thought experiment involves a demon controlling a trapdoor between two gas chambers, allowing fast particles to pass into one chamber and slow particles into the other. Using *information* about the velocity of the particles, the demon creates a temperature difference, reducing the entropy without expending energy (see Fig. 1 for a pictorial representation and historical context).

The demon's ability to reduce entropy without performing work hinges entirely on the information it possesses about the gas particles. This suggests that information somehow plays an important role in characterizing thermodynamic processes. This interplay was tightened with the Szilárd engine [14,15]—a more operational version of Maxwell's demon, based on a device that trades information for work. The setup consists of a chamber containing a single gas molecule in thermal equilibrium with a heat bath at temperature T [16]. Based on information about which half the chamber of the molecule is in, a thin wall attached to a weight is inserted. The one-molecule gas then pushes the wall, expanding the chamber back to its initial size and extracting $k_B T \log 2$ of work from the thermal reservoir (approximately 3×10^{-21} J at room temperature). Like Maxwell's demon, the Szilárd engine appears to violate the second law of thermodynamics, as the heat extracted from the environment is entirely converted into mechanical work. Although Szilárd acknowledged the need for an entropy increase to close the thermodynamic cycle, he did not specify its source—whether the increase arose from measuring the particle's position, storing the collected information, inserting and removing the partition, or erasing the recorded data.

In the following years, information and computation theories were developed [17–19]. Although they were not yet tied to physics, Shannon's theory [18] inspired Brillouin [20] to attempt to save the second law. He proposed a mathematical theory that linked the acquisition of information and the increase in entropy needed to restore the second law. Two years after Brillouin's paper, Landauer [21,22] asked whether thermodynamics imposes physical

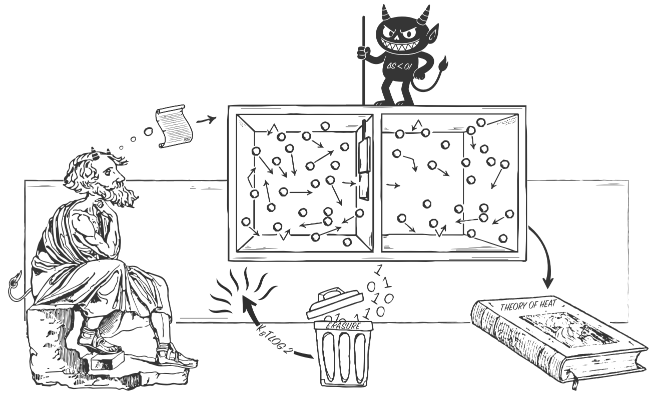


FIG. 1. *Maxwell's Gedanken experiment*. In 1867 James Clerk Maxwell addressed a letter to Peter Guthier Tait with the provocative idea that the second law is a statistical principle that only holds on average [13]. There, he introduced his famous demon—an intelligent being capable of sorting gas molecules by their velocities. By creating a temperature difference, the demon apparently defies the second law of thermodynamics, decreasing the system's entropy without expending energy. Later on, this apparent paradox appeared in his book "*Theory of Heat*" [12].

limits on information processing. The answer to such a simple question led to the conclusion that erasing information is inherently linked to a heat cost. Finally, this intriguing connection between thermodynamics and computation was later strengthened by Roger Penrose [23] and Charles Bennett [24], which ultimately led to the resolution of Maxwell's demon [25].

What seemed to be the end of the demon's journey turned out to be just the beginning. Both stochastic thermodynamics [26] and quantum thermodynamics [27–33] have reinterpreted and extended the concept of Maxwell's demon beyond its original scenario. Beautifully, not only has theoretical progress been made, but numerous experiments on these two topics have also been reported (a summary of experiments in the field of stochastic thermodynamics can be found in Ref. [34] and an overview of the proposed and realized quantum thermodynamic devices is presented in Ref. [35]).

How to use this guide?

This tutorial revisits Maxwell's demon, from its early phase to modern approaches, while providing historical context and connections to state-of-the-art experiments. It offers an introduction for quantum scientists, exploring the intersection of thermodynamics, information theory, and quantum mechanics. We do not intend to provide a detailed review of the literature here; for that, we recommend both the reviews [36–38] and the book by Leff and Rex [39]. In each section, we discuss problems (along with their solutions) relevant to the topic. These problems often introduce results that will be useful in subsequent discussions. To highlight these key results, they are marked with the symbol ✎ and are enclosed between horizontal separators. Starting from Sec. IV, we present experimental realizations related to the discussed topic. These realizations are independent and can be skipped without compromising the understanding of the main text. The experiments are placed in boxes, and the discussion is kept short. Our goal is to show that what might seem theoretical has already been demonstrated experimentally. The experiments were selected based on the physical platforms rather than their chronological order of appearance in the literature.

This tutorial is organized as follows. We start with a brief recap of basic thermodynamics concepts in Sec. II, followed by a concise summary of relevant quantum mechanics notions in Sec. III. The central problem of this tutorial is presented in Sec. IV, where Maxwell's thought experiment is discussed (Sec. IV A), along with Szilard's reinterpretation (Sec. IV B) and attempts to resolve the apparent violation of the second law of thermodynamics (Sec. IV C). Next, Sec. V starts with a discussion of what information is, introducing Shannon's theory and some mathematical results (Sec. V A). Then we briefly touch on some basic ideas and historical facts about computation (Sec. V C). One of the most important points of this tutorial is then presented in Sec. VI, where Landauer's principle is introduced (Sec. VI A), rigorously proven (Sec. VI B), and its first experimental verification is discussed (Sec. VI C). To finish this section, the notion of reversible computation is briefly explained (Sec. VI D). The resolution of Maxwell's demon is presented in Sec. VII and three different approaches to solving this "paradox" are discussed. Finally, in Sec. VIII, we briefly discuss recent progress related to Maxwell's demon, as well as topics inspired by it.

II. THERMODYNAMICS IN A NUTSHELL

This introductory section revisits some fundamental formulations of thermodynamics, introduces the notation used, and briefly highlights recent developments in the field. A few problems, which will later connect to other sections, are also discussed. As this section is a warmup, readers already familiar with thermodynamics are welcome to skip it.

Thermodynamics emerged from the desire to understand how macroscopic systems exchange energy, particularly the interconversion between heat and work [40,41]. It offers a framework for describing the transformations of complex systems made up of many particles without the need to take into account their microscopic details. The key insight is that macroscopic systems in equilibrium can be very well described by just a few variables, like temperature, pressure, or magnetization, depending on the specific physical system being studied. This allows us to make accurate predictions about the system behavior without having to track every individual particle. While this might seem like an oversimplification, the universality of thermodynamic principles led to the formulation of four fundamental laws [42]:

(0) **Zeroth law.** *If two systems are each in thermal equilibrium with a third system, they are also in thermal equilibrium with one another.*

(1) **First law.** *Energy is conserved. For a system interacting with an environment, energy transfer can be divided into two contributions: work and heat.*

(2) **Second law.** *The entropy of an isolated system undergoing a physical process can never decrease.*

(3) **Third law.** *It is impossible to attain absolute zero in a finite number of thermodynamic operations. At absolute zero, systems with degenerate ground states may retain nonzero entropy.*

The central focus of this tutorial is on the second law of thermodynamics, which can be formulated in various ways, with all formulations being fundamentally equivalent. In its most general form, the second law asserts that the entropy of the universe tends to increase $\Delta S_{\mathsf{univ}} \geq 0$. The change in the system's entropy, known as entropy production, is denoted by $\Sigma := \Delta S_{\mathsf{univ}}$. Thus, a succinct way to express the second law is through the non-negativity of the entropy production. This quantity also allows us to classify processes as reversible, where $\Sigma = 0$, or irreversible, with $\Sigma > 0$. For a detailed discussion of entropy production in both classical and quantum systems, see Ref. [43].

Throughout this tutorial, a main system interacts with its environment, resulting in changes to both the energy and the entropy of each. The total entropy of the universe is divided into the entropy of the system and that of the environment: $S_{\text{univ}} = S_{\mathsf{S}} + S_{\mathsf{B}}$. This leads to the second law being written as $\Delta S_{\text{univ}} = \Delta S_{\mathsf{S}} + \Delta S_{\mathsf{B}} \geq 0$. Moreover, we assume that the environment consists of a work source, which exchanges only energy with the system, leaving its entropy unchanged, and a heat bath, which exchanges both energy and entropy with the system. Thus, ΔS_{B} simplifies to the change in entropy of the heat bath, which is modeled as being in equilibrium at temperature T. The change in entropy of the heat bath is then given by $\Delta S_{\mathsf{B}} = \int \mathrm{d}Q/T$, where $\mathrm{d}Q$ represents an infinitesimal heat flow into the system. The total heat exchanged is $Q = \int \mathrm{d}Q$. Here, we adopt the convention that heat is positive when it flows into the system from the environment. Work, however, is performed by an external controlling agent—an entity that manipulates macroscopic parameters (e.g., volume, magnetic fields) to exchange energy with the system. Crucially, the agent's ability to extract or inject work depends on what they know about the system's state. In typical thermodynamics, the agent possesses only macroscopic information (temperature, volume, pressure, etc). However, a hypothetical all-knowing agent—like the star of this tutorial, the "famous" Maxwell's demon—would be privy to microscopic information about each molecule. Such an agent could, in principle, exploit thermal fluctuations to convert heat entirely into work, seemingly violating the second law. As we will discuss in detail later (via Landauer's principle), this paradox is resolved when one accounts for the entropy production associated with the demon's measurement process and the erasure of its memory, which compensates for any apparent entropy decrease in the system. For now, we retain the classical assumption of macroscopic control and minimal knowledge, leading us to the second law in its traditional form, Clausius's inequality [40]:

$$\Sigma = \Delta S_{\mathrm{S}} + \int \frac{\mathrm{d}Q}{T} \geq 0. \tag{1}$$

If the heat bath is only slightly perturbed, meaning it is large enough so that its temperature remains effectively constant, the second term in Eq. (1) simplifies to Q/T. From this point onwards, we assume that the bath remains at its initial temperature at all times and that the system does not significantly perturb it. In the presence of multiple heat baths Eq. (1) generalizes naturally to $\Sigma = \Delta S_{\mathrm{S}} + \sum_i Q_i/T_i \geq 0$, where the index i refers to the ith reservoir.

Traditionally, classical thermodynamics describes quasistatic (reversible) transformations, in which the system and every reservoir remain in equilibrium at each instant. In that equilibrium limit, the entropy-production rate vanishes, $\dot{\Sigma} = 0$, and, because no heat or particle currents flow, the entropy flux—defined as $\dot{\Phi} := \sum_i \dot{Q}_i/T_i$—is also zero $\dot{\Phi} = 0$, so that $\mathrm{d}S_{\mathrm{S}}/\mathrm{d}t = 0$ (where t is time). However, many processes occur over finite times and lead to entropy production. A scenario that can emerge after finite-time transients is one where there is a continuous flow of energy or matter, with certain system properties becoming constant in time despite the absence of thermal equilibrium. After this transient period has elapsed, the condition $\dot{S}_{\mathrm{S}} = 0$ together with the balance relation $\dot{S}_{\mathrm{S}} = \dot{\Sigma} - \dot{\Phi}$ implies $\dot{\Sigma} = \dot{\Phi} > 0$, which means that entropy is continually generated within the system and is transferred entirely to the reservoirs. This defines a nonequilibrium steady state [44], in which the entropy-production rate remains constant.

The concepts presented so far can be illustrated by considering a thermodynamic system interacting with heat baths at different temperatures and a work reservoir. The notion of entropy production allows us to incorporate time into the analysis and describe the continuous interaction between the system and its environment. Consequently, we express the first and second laws in terms of rates, as follows:

$$\frac{\mathrm{d}E_{\mathrm{S}}}{\mathrm{d}t} = \dot{W} - \sum_i \dot{Q}_i, \tag{2}$$

$$\frac{\mathrm{d}S_{\mathrm{S}}}{\mathrm{d}t} = \dot{\Sigma} - \sum_i \frac{\dot{Q}_i}{T_i}. \tag{3}$$

where $\dot{E}_{\mathrm{S}}$ represents the rate of change of the system's internal energy, $\dot{Q}_i$ denote the heat exchanged with ith bath, respectively, and $\dot{W}$ is the rate of work done on the system, which is assumed to be positive in the case of work being supplied to the system.

Let us discuss three paradigmatic examples (presented in Ref. [43]), in which we analyze a system that has operated for a sufficiently long time and has reached a steady state. This implies that both Eqs. (2) and (3) are equal to zero, and the steady state is therefore characterized by a continuous conversion of heat into work, accompanied by a constant production of entropy.

Suppose that there is no work involved in the process and only heat exchange occurs between the two thermal reservoirs at temperatures T_c and T_h with $T_c < T_h$. The heat exchanged with the hotter and colder reservoirs are denoted by $\dot{Q}_h$ and $\dot{Q}_c$, respectively. As a result, the first law simplifies to $\dot{Q}_h = -\dot{Q}_c$, which, when substituted into Eq. (3), gives

$$\dot{\Sigma} = \left(\frac{1}{T_c} - \frac{1}{T_h}\right)\dot{Q}_c \geq 0. \tag{4}$$

Since the above equation must be non-negative and $T_c < T_h$, it follows directly that $\dot{Q}_c \geq 0$, which according to our convention means that heat must flow from the hot to the cold reservoir. This is precisely Clausius's statement of the second law:

"A transformation whose only final result is to transfer heat from a body at a given temperature to a body at a higher temperature is impossible." [45]

Now, consider the case where the system is connected only to the hot reservoir. In this case, the first law becomes $\dot{W} = \dot{Q}_h$, and the second law is expressed as

$$\dot{\Sigma} = \frac{\dot{Q}_h}{T_h} = \frac{\dot{W}}{T_h} \geq 0. \tag{5}$$

By definition, positive work implies that work is done on the system rather than extracted from it. Therefore, work cannot be extracted from a single heat reservoir, which directly corresponds to Kelvin's statement:

"A transformation whose only final result is to transform into work heat extracted from a source which is at the same temperature throughout is impossible." [46]

As a final example, we now consider a machine comprised of two heat baths at temperatures T_c and T_h, and a work reservoir. Assuming that this machine has reached a steady state allows us to rewrite Eqs. (2) and (3) as $\dot{W} = \dot{Q}_h + \dot{Q}_c$ and $\dot{\Sigma} = \dot{Q}_h/T_h + \dot{Q}_c/T_c$. Computing the efficiency of the engine (i.e., the ratio of output work to input heat), we arrive at

$$\eta = \frac{\dot{W}}{\dot{Q}_h} = 1 + \frac{\dot{Q}_c}{\dot{Q}_h} = \underbrace{\left(1 - \frac{T_c}{T_h}\right)}_{:=\eta_C} + \frac{T_c}{\dot{Q}_h}\dot{\Sigma}. \tag{6}$$

The term in brackets represents the well-known Carnot efficiency η_C. Since $\dot{Q}_h < 0$, as the heat flows from the hot reservoir to the system, the second term is strictly negative. Therefore, the efficiency of a heat engine operating in finite time is given by Carnot's efficiency with a negative correction proportional to the entropy production. This result leads to Carnot's statement of the second law:

"The efficiency of any reversible engine operating between two heat reservoirs depends only on the reservoir temperatures and is equal to the Carnot efficiency. No heat engine operating between the same two temperatures can be more efficient." [47]

The previous example shows that incorporating the concept of entropy production provides a deeper understanding of Carnot's theorem. Specifically, the efficiency of a heat engine operating in finite time can always be expressed as the Carnot efficiency minus a correction term, directly related to the irreversibility of the process.

So far, we have made general statements about thermodynamic processes, mostly for systems in thermal equilibrium, but also considering nonequilibrium states. However, we have remained completely agnostic about the "form" of the system's state. Since the Gibbs state is crucial for much of what follows, we will now simply present its expression without justifying its form—readers are instead referred to the books by Landau and Lifshitz [48], Pathria [49], or Callen [41] for a more detailed explanation. To do so, we consider a classical continuous system [50] in contact with a heat bath at temperature T. The system is described by a Hamiltonian $H(z, \lambda)$, where $z = (q, p)$ represents a point in the phase space, and λ is a real parameter (e.g., volume, magnetic field, or any other externally controllable parameter). We then define a thermal state as a probability distribution over phase space

$$\rho_{\mathsf{E}}(z, \lambda) := \frac{e^{-\beta H(z,\lambda)}}{Z(\lambda)}, \tag{7}$$

where $Z(\lambda) = \int \mathrm{d}z\, e^{-\beta H(z,\lambda)}$ is the partition function.

Average thermodynamic quantities

Using Eq. (7), one can calculate important average quantities for the system. For example, the average equilibrium energy is given by

$$\begin{aligned} E_{\mathsf{S}}(\lambda) &:= \int \mathrm{d}z \frac{e^{-\beta H(z,\lambda)}}{Z(\lambda)} H(z, \lambda) = \frac{\frac{-\partial}{\partial\beta}\int \mathrm{d}z e^{-\beta H(z,\lambda)}}{Z(\lambda)} \\ &= -\frac{\partial}{\partial\beta}\log Z(\lambda). \end{aligned} \tag{8}$$

Similarly, the Gibbs entropy takes the form (again, yet without justifying the reasoning behind this equation):

$$\begin{aligned} \frac{S(\lambda)}{\beta} &= -\frac{1}{\beta}\int \mathrm{d}z \frac{e^{-\beta H(z,\lambda)}}{Z(\lambda)} \log\left(\frac{e^{-\beta H(z,\lambda)}}{Z(\lambda)}\right) \\ &= E_{\mathsf{S}}(\lambda) + \frac{\log Z(\lambda)}{\beta}. \end{aligned} \tag{9}$$

Another important formulation of the second law is given in terms of free energy F. When thermodynamic transformations are described between equilibrium states at constant temperature (isothermal processes), this quantity provides a bound on the work that can be extracted in the process. Namely,

$$\langle W \rangle \geq \Delta F. \tag{10}$$

Here, $F := E - TS$, where E and S denote the system's internal energy and thermodynamic entropy, respectively. To simplify notation, we omit the index S when the context is clear and include it only when needed to avoid confusion. Note that from Eq. (9), we directly get $F(\lambda) = -(\partial/\partial\beta)Z(\lambda)$. Moreover, we can verify Eq. (10) considering a simple protocol as in the example below:

Free energy bound

Consider a classical system whose microstate is $z = (q, p)$ and whose Hamiltonian depends on a control parameter $\lambda \in [0, 1]$. The protocol lasts a time τ and we take a constant switching rate $\dot{\lambda} = 1/\tau$. For a trajectory that begins in z_0 and follows Hamiltonian dynamics, the mechanical work is

$$W_{z_0} = \int_0^\tau \mathrm{d}t\dot{\lambda}\frac{\partial H(z(t),\lambda)}{\partial\lambda} = \int_0^1 \mathrm{d}\lambda\,\frac{\partial H(z(\lambda),\lambda)}{\partial\lambda}. \quad (11)$$

Averaging Eq. (11) over initial conditions drawn from $\rho(z_0)$ gives

$$\langle W\rangle = \int \mathrm{d}z_0\rho(z_0)W_{z_0}, \quad (12)$$

where $\rho(z_0)$ is the probability distribution of the initial state, which can be in or out of equilibrium. For an equilibrium system, this reduces to the Boltzmann distribution, as given in Eq. (7), leading to

$$\langle W\rangle = \int_0^1 \mathrm{d}\lambda \int \mathrm{d}z\frac{e^{-\beta H(z,\lambda)}}{Z(\lambda)}\frac{\partial H(z,\lambda)}{\partial\lambda}. \quad (13)$$

In the limit $\tau \to \infty$, the system remains in instantaneous equilibrium, and Eq. (13) reduces to

$$\begin{aligned}\langle W\rangle &= -\frac{1}{\beta}\int_0^1 \frac{\mathrm{d}}{\mathrm{d}\lambda}\ln Z(\lambda)\,\mathrm{d}\lambda = -\frac{1}{\beta}[\ln Z(1) - \ln Z(0)]\\ &= F(1) - F(0), \end{aligned} \quad (14)$$

where $F(\lambda) = -\beta^{-1}\ln Z(\lambda)$ was noted below Eq. (9). Therefore, a quasistatic isothermal transformation extracts an average work $\langle W\rangle = \Delta F$ saturating the free-energy bound Eq. (10). Any finite-time drive ($\tau < \infty$) produces additional dissipation, so that $\langle W\rangle > \Delta F$.

The last example shows that, for an infinitely slow process, the work done on the system is exactly equal to the free energy difference between the final and initial states. Conversely, finite-time processes lead to irreversibility, and the thermodynamic work required to transform a system, in contact with an environment, can be expressed as $W = \Delta F + W_{\mathrm{diss}}$, where the dissipated work is directly proportional to the entropy production [43,51].

Surprisingly, the inequality corresponding to the second law as given in Eq. (10) can be expressed as an equality [52,53], namely [54],

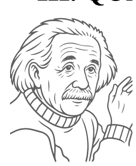
$$\langle e^{-\beta W}\rangle = e^{-\beta\Delta F}, \quad (15)$$

where $\langle\cdot\rangle$ denotes an average over multiple realizations of the given protocol, and $\beta = 1/k_BT$. This expression, known as the Jarzynski equality, holds for arbitrary processes, even those far from equilibrium. It is remarkable because it connects equilibrium and nonequilibrium quantities: while the left-hand side represents the averaged work over many realizations of a process, the right-hand side corresponds to the free energy difference between two equilibrium states. Its applicability extends beyond physics [55].

Understanding the thermodynamics of systems far from equilibrium, where fluctuations cannot be neglected, is anything but simple. However, if the system evolves under well-defined external protocols—for instance, via external driving that deterministically changes the parameters describing the system—thermodynamic quantities such as work, heat, or entropy production can be treated as random variables, each characterized by a probability distribution constructed over many realizations of the protocol. Remarkably, fluctuations in these quantities satisfy universal constraints known as fluctuation theorems [56,57], with the Jarzynski equality being a prominent example of an integral fluctuation theorem. While we do not explore the full details of fluctuation theorems in this tutorial, it is important to note that many of the concepts and illustrations we discuss here are based on and derived from these foundational relations.

Originally, free energy was defined only for states in thermal equilibrium. However, given its operational meaning, the definition can be extended to nonequilibrium states [58,59]. Specifically, in the next section, we will naively replace the thermodynamic entropy with a different notion of entropy. This will allow us to define a *nonequilibrium free energy*. In this broader context, the system interacts with a thermal reservoir at temperature T, although the system itself may not have a well-defined temperature. While we will not explicitly demonstrate it here, nonequilibrium free energy quantifies the maximum average work that can be extracted from a system in an out-of-equilibrium state [60,61].

III. QUANTUM MECHANICS IN A NUTSHELL



"Quantum mechanics: Real Black Magic Calculus." [62]

For completeness and to make this tutorial as self-contained as possible, in this section, we will cover the basic tools needed to describe quantum systems, focussing on what is useful to discuss Maxwell's problem and information erasure later on. As a result, we avoid going into too much detail and are less rigorous in both notation and justification—proving Einstein's words to be surprisingly accurate. For those who want a comprehensive and detailed material, we recommend the books by Sakurai [63], Peres [64], Nielsen and Chuang [65], or Bertlmann and Friis [66]. Readers already familiar with quantum mechanics and quantum information theory are welcome to jump ahead—it will not make resolving Maxwell's demon any more difficult.

To simplify the mathematical description, we limit our discussion to finite-dimensional systems, as many experimentally relevant physical systems can be well approximated this way. To keep the discussion general, we avoid focussing on specific physical systems while still aiming to keep the explanations as simple as possible. This approach allows us to focus on the core ideas without getting bogged down in unnecessary complexity.

In a classical system with a configuration space of size $d < \infty$, the system can exist only in one of the d distinct configurations. However, in quantum theory, describing a system with d configurations means associating it with a finite-dimensional Hilbert space $\mathcal{H} \in \mathbb{C}^d$, where each configuration corresponds to a vector in an orthonormal basis. But what exactly is a Hilbert space and how do we construct one?

For a system with d distinct configurations labeled $\{1, \dots, d\}$, the associated Hilbert space $\mathcal{H}$ is a complex vector space equipped with an inner product. Each configuration corresponds to a vector $|i\rangle \in \mathcal{H}$, and the set $\{|1\rangle, |2\rangle, \dots, |d\rangle\}$ forms an orthonormal basis for $\mathcal{H}$. Our workhorse in this tutorial is the quantum version of a classical bit, the qubit, which can be described by two configurations, 0 and 1. The Hilbert space of the qubit is $\mathcal{H} = \mathbb{C}^2$, and the two classical configurations correspond to the quantum states $|0\rangle$ and $|1\rangle$ in $\mathcal{H}$. Examples of such systems include a photon of light with two orthogonal polarizations, an electron with two possible spin states, or a two-level atom with two possible energy states, among many others.

We can move on to the postulates of quantum mechanics, which will guide us in (i) defining the possible states of a quantum system, (ii) understanding and calculating measurement outcomes, and (iii) describing how the system evolves over time.

(I) A pure state of a quantum system is associated with a vector in the Hilbert space, represented by $|\psi\rangle \in \mathcal{H}$.

In some cases, we can only associate a vector $|\psi_i\rangle$ with a classical probability p_i. For example, imagine placing qubits randomly in a box, sometimes in the state $|0\rangle$ and sometimes in the state $|1\rangle$. If we blindly pick a qubit from the box, the state will be $|0\rangle$ or $|1\rangle$, each with a probability of $\frac{1}{2}$. These states are called mixed states and are mathematically described by an operator $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. As a result, the previous postulate can be refined as follows:

(I') To any state of a system, we associate a density operator ρ, which (i) is Hermitian ($\rho = \rho^\dagger$), (ii) is positive semidefinite ($\rho \geq 0$), and (iii) has a trace equal to one [$\mathrm{tr}(\rho) = 1$].

(II) An observable is associated with a Hermitian operator $O = O^\dagger$ that acts on the Hilbert space of the system.

Given an observable O, we can always express its spectral decomposition as $O = \sum_i o_i \mathbb{P}_i$, where o_i are real numbers, and $\mathbb{P}_i$ are orthogonal projection operators onto the eigenspace corresponding to the eigenvalue o_i. These projectors satisfy $\mathbb{P}_i \mathbb{P}_{i'} = \delta_{i,i'} \mathbb{P}_i$ and $\sum_i \mathbb{P}_i = \mathbb{1}$.

(III) Given a state ρ, when we measure an observable O, the outcome will be one of its eigenvalues o_i. The probability p_i of obtaining this outcome and the state ρ_i after the measurement are given by $p_i = \mathrm{tr}(\mathbb{P}_i \rho)$ and $\rho_i = \mathbb{P}_i \rho \mathbb{P}_i / p_i$, respectively.

Note that in quantum measurements, the act of measuring generally changes the state of the system, typically collapsing it into one of the eigenstates of the observable [67]. This postulate also tells us that it is possible for the same experiment, repeated under identical conditions, to give different outcomes.

(IV) The dynamics of a closed system is defined by a unitary operator U that satisfies $UU^\dagger = U^\dagger U = \mathbb{1}$ as

$$|\psi\rangle \to |\psi'\rangle = U|\psi\rangle \quad \text{and} \quad \rho \to \rho' = U\rho U^\dagger. \tag{16}$$

In many cases of interest, we can express the time-evolution operator as $U = e^{-iHt}$, where H is the Hamiltonian, t is the elapsed time, and we set $\hbar = 1$. Importantly, unitary evolution occurs when the system is completely isolated, which we refer to as a closed system. However, systems are rarely fully isolated, as they typically interact with other degrees of freedom, known as the environment. Accounting for the environment introduces the possibility of more general dynamics [68–70], which will also not be directly discussed in this tutorial, although many examples discussed here are described in open dynamics.

Let us now briefly comment on a fundamental quantity that appears in von Neumann's seminal work from 1932 [71], which today bears his name. Motivated by a desire to better understand the nature of the measurement process, he devised a thought experiment [72] (this time without challenging the second law) and ended up deriving what we now call the von Neumann entropy:

$$S(\rho) := -\mathrm{tr}(\rho \log \rho) = -\sum_i p_i \log p_i, \tag{17}$$

where p_i are the eigenvalues of ρ. The equation above quantifies the amount of mixedness or lack of knowledge about the exact state of a system when it is described by a density matrix ρ. Therefore, we often say that the von Neumann entropy is a measure of the system's mixedness. Consequently, the entropy of a pure state is zero because there is no ambiguity about the system's state. Mathematically, this is evident since a pure state has one eigenvalue $\lambda_1 = 1$, with all the others zero. However, if the system is in a mixed state, its entropy is larger than zero, which captures our incomplete knowledge about the system. In the extreme case, where we have no information about the system's state (i.e., when it is maximally mixed), its entropy reaches the maximum value of $\log d$, where d is the dimension of the system.

The von Neumann entropy has many interesting properties. In the upcoming discussion, we will focus on three relevant ones. Although these can be proven formally, we will skip the proofs and instead rely on their physical interpretations to understand the results. The first property is that von Neumann entropy is additive for uncorrelated systems: given ρ_{A} and ρ_{B}, describing independent systems A and B, we have

$$S(\rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}}) = S(\rho_{\mathsf{A}}) + S(\rho_{\mathsf{B}}). \tag{18}$$

This property simply tells us that in the absence of any correlation between A and B, the total uncertainty is just the sum of the uncertainties of the individual systems. The second property is that the von Neumann entropy is invariant under unitary operations

$$S(\rho) = S(U\rho U^{\dagger}). \tag{19}$$

Unitary evolution corresponds to a change of basis in the Hilbert space. Since no information is lost and unitary processes are reversible, it is expected that the von Neumann entropy remains unchanged in such cases. Finally, the von Neumann entropy is subadditive. Given two subsystems A and B, the entropy of the composite system satisfies the following inequality:

$$S(\rho_{\mathsf{AB}}) \leq S(\rho_{\mathsf{A}}) + S(\rho_{\mathsf{B}}). \tag{20}$$

This inequality captures the fact that there may be correlations between the two subsystems, which can reduce the total uncertainty about the joint system. Consequently, it is saturated only when the two systems are uncorrelated. Thus, we can use this as a measure of the total degree of correlations by defining the mutual information:

$$I(\mathsf{A}:\mathsf{B})_{\rho_{\mathsf{AB}}} := S(\rho_{\mathsf{A}}) + S(\rho_{\mathsf{B}}) - S(\rho_{\mathsf{AB}}). \tag{21}$$

This quantity represents the amount of information stored in AB that is not contained in A and B individually. However, it is important to note that mutual information quantifies the total degree of correlations without distinguishing between classical and quantum contributions. Another important quantity that we will come across in the next sections is the quantum relative entropy, or the Kullback-Leibler divergence [73]. Given two density matrices σ and ρ, it is defined as

$$S(\sigma\|\rho) := \mathrm{tr}[\sigma(\log\rho - \log\sigma)] = -S(\sigma) + \mathrm{tr}(\sigma\log\rho). \tag{22}$$

This quantity is important to us because it satisfies the Klein inequality $S(\sigma\|\rho) \geq 0$, and $S(\sigma\|\rho) = 0$ if and only if $\sigma = \rho$.

In Sec. V, we will introduce another entropy, which is mathematically very similar to the von Neumann entropy and can be interpreted as its classical counterpart. Importantly, while von Neumann and thermodynamic entropy frequently coincide—particularly for equilibrium states described by Gibbs ensembles—this is not guaranteed in general. A simple example illustrating this discrepancy is a single isolated quantum system prepared in a pure energy eigenstate. Suppose a system with Hamiltonian H is prepared in a pure state $|E\rangle$ of energy E. Its density operator is $\rho = |E\rangle\langle E|$, which has zero von Neumann entropy. However, if the system were in thermodynamic equilibrium at the same energy E, we would describe it using a canonical or microcanonical ensemble, each typically possessing nonzero thermodynamic entropy. This contrast demonstrates that the two notions of entropy cannot, in general, be identified—similar differences can arise in other contexts, such as strong coupling [74–76] or nonequilibrium scenarios [77,78]. We will return to this discussion at the end of Sec. V.

✎ *Entropy of a system at thermal equilibrium*

Consider a system, described by a Hamiltonian H, and prepared in a thermal Gibbs state at inverse temperature β:

$$\gamma = \frac{e^{-\beta H}}{Z} \quad \text{where } Z := \mathrm{tr}(e^{-\beta H}). \tag{23}$$

Substituting Eq. (23) into Eq. (17), we obtain

$$\begin{aligned} S(\gamma) &= -\mathrm{tr}(\gamma\log\gamma) = -\mathrm{tr}(\gamma\log\frac{e^{-\beta H}}{Z}) \\ &= -\mathrm{tr}[\gamma\,(-\beta H - \mathbb{1}\log Z)] \\ &= \beta\mathrm{tr}(\gamma H) + \log Z. \end{aligned} \tag{24}$$

✎

By defining the average energy as $E := \mathrm{tr}(\gamma H)$ and identifying the last term as the free energy $F := -\log Z/\beta$ in the canonical ensemble [49], where $Z = \mathrm{tr}(e^{-\beta H})$, the free energy of the system can be expressed in terms of the von Neumann entropy as $F = E - \beta^{-1}S(\gamma)$.

Motivated by its operational role as the maximum average extractable work, the free energy can be extended from equilibrium to arbitrary quantum states. Consider a finite-dimensional system with Hamiltonian H in an arbitrary state ρ, whose average energy is $E(\rho) := \mathrm{tr}(\rho H)$. If the system exchanges energy with a heat bath at inverse temperature β, we define the nonequilibrium free energy [58,59]

$$\mathcal{F}(\rho) := E(\rho) - \frac{1}{\beta}S(\rho), \\ \stackrel{(22)}{=} \frac{1}{\beta}[S(\rho\|\gamma) - \log Z], \tag{25}$$

where $S(\rho)$ is the von Neumann entropy given by Eq. (17), γ is thermal Gibbs state defined in Eq (60) and Z is the partition function. Because $S(\rho\|\gamma) \geq 0$ with the equality if and only if $\rho = \gamma$, the Gibbs state uniquely minimizes $\mathcal{F}(\rho)$. Equivalently, writing the equilibrium free energy $F = -\beta^{-1}\log Z$, Eq. (25) can be recast as $\mathcal{F} = \beta^{-1}S(\rho\|\gamma) + F$.

✎ *Recording quantum information*

As a warmup example for what follows, imagine that we have a particle in a box, where its exact position is unknown, but we know it can be in one of two possible states: either left (L) or right (R). We will demonstrate that by using an additional subsystem, a *memory*, we can store the position of the particle in the memory. To describe this scenario, we first assume the particle is in a mixed state while the memory is in a blank state, so the composite system is represented by the following density operator:

$$\rho_{\mathsf{SM}} = \frac{(|L\rangle\langle L|_{\mathsf{S}} + |R\rangle\langle R|_{\mathsf{S}})}{2} \otimes |0\rangle\langle 0|_{\mathsf{M}}\,. \tag{26}$$

We now apply a controlled-NOT (CNOT) gate in the composite system. This is a unitary operation that acts as follows: $U_{\mathrm{CNOT}} = |L\rangle\langle L|_{\mathsf{S}} \otimes \mathbb{1} + |R\rangle\langle R|_{\mathsf{S}} \otimes \sigma_x$, where σ_x is the Pauli-x matrix. Its effect can be understood by observing how it transforms the states:

$$|L\rangle_{\mathsf{S}} \otimes |0\rangle_{\mathsf{M}} \to |L\rangle_{\mathsf{S}} \otimes |0\rangle_{\mathsf{M}}\,, \\ |R\rangle_{\mathsf{S}} \otimes |0\rangle_{\mathsf{M}} \to |R\rangle_{\mathsf{S}} \otimes |1\rangle_{\mathsf{M}}\,. \tag{27}$$

As we can see, this unitary operation changes the state of the second qubit conditioned on the state of the first one. Hence, after the unitary, the composite state of the system is now given by

$$\sigma_{\mathsf{SM}} = U_{\mathrm{CNOT}}\rho_{\mathsf{SM}}U_{\mathrm{CNOT}}^{\dagger} = \frac{1}{2}(|L,0\rangle\langle L,0| + |R,1\rangle\langle R,1|). \tag{28}$$

We observe that if the particle is on the left, the memory remains in the blank state $|0\rangle_{\mathsf{M}}$, but if the particle is on the right, the memory is in the state $|1\rangle_{\mathsf{M}}$. Once we know the particle's position, we can proceed to use this information for a thermodynamic protocol or any other information-processing task that we wish. However, note that if we now look at the state of the memory [which corresponds to tracing out the main system in Eq. (28)], we find that the memory is in a mixed state. However, if we measure the state of the memory, we will obtain a pure state for the system, conditioned on the outcome of the measurement. This means that the memory is entangled with the system, and the memory can be used to extract information about the system. In practice, if we discard the measurement result (i.e., ignore the outcome), the memory is in a mixed state. In this case, the memory can no longer be reused for the same process unless it is reset to its initial state or replaced with a new one.

✎

Let us summarize this example. We begin with the system in an unknown state, while the memory is in a blank known state. After the operation, the memory becomes correlated with the system: it is in the state $|0\rangle_{\mathsf{M}}$ if the particle is on the left side of the box $|L\rangle_{\mathsf{S}}$, or in the state $|1\rangle_{\mathsf{M}}$ if the particle is on the right side $|R\rangle_{\mathsf{S}}$. This information can then be used to perform a specific protocol, such that by the end of the transformation, the system returns to its initial state, but the memory does not—it ends up in a mixed state between blank and nonblank. We will revisit this problem in Sec. VII and reinterpret it in light of its connection to Maxwell's demon.

✎ *Simple model for thermodynamic work in quantum mechanics*

Consider a closed system described by a Hamiltonian H, satisfying the eigenproblem $H|\psi_n\rangle = \epsilon_n|\psi_n\rangle$, where $|\psi_n\rangle$ and ϵ_n denote the nth eigenstate and eigenenergy, respectively. The average energy of the system is given by $E = \sum_n p_n\epsilon_n$, where p_n represents the occupation probability of the nth eigenstate. Note that if the system is in thermal equilibrium, p_n is determined by the Gibbs distribution [Eq. (60)]. The differential of the internal energy E is

$$\mathrm{d}E = \sum_n(\underbrace{\epsilon_n\mathrm{d}p_n}_{\mathrm{d}Q} + \underbrace{p_n\mathrm{d}\epsilon_n}_{-\mathrm{d}W}). \tag{29}$$

From the first law of thermodynamics for an isothermal process, we have $T\mathrm{d}S = \mathrm{d}E + \mathrm{d}W$. Consequently, we can identify that the first term in Eq. (29) is the heat exchanged, and the second term is the negative of the work done by the system. This decomposition is not universal; rather, it is motivated by scenarios where the system's Hamiltonian depends on an external parameter λ that can be modified, thus changing ϵ_n—for example, by applying external fields. In this context, the second term should be associated with $T\mathrm{d}S$, as the entropy is given by Eq. (17). Let us now derive an expression for thermodynamic work in an isothermal process when the system starts at thermal equilibrium at inverse temperature β and the external parameter λ is varied from λ_1 to λ_2 at constant temperature. First, we note the following identity:

$$-\frac{1}{\beta}\frac{\partial \log Z}{\partial \epsilon_m} = -\frac{1}{\beta Z}\frac{\partial Z}{\partial \epsilon_m} = \frac{1}{Z}\left[\frac{\partial}{\partial \epsilon_m}\sum_n e^{-\beta\epsilon_n}\right] = \frac{e^{-\beta\epsilon_n}}{Z} = p_n. \tag{30}$$

Since ϵ_n depends on λ, substituting Eq. (30) in the expression for W, we obtain

$$W = \frac{1}{\beta}\sum_n \int_{\lambda_1}^{\lambda_2}\left(\frac{\partial \log Z}{\partial \epsilon_n}\right)\left(\frac{\partial \epsilon_n}{\partial \lambda}\right) \mathrm{d}\lambda = \frac{1}{\beta}[\log Z(\lambda_2) - \log Z(\lambda_1)]. \tag{31}$$

This expression relates the work done during an isothermal process, as the external parameter λ varies from λ_1 to λ_2.

✎

In the previous example, we introduced a simple model for thermodynamic work in quantum mechanics. In general, the definition of work in quantum thermodynamics depends on the specific context (see Ref. [32] for a detailed discussion), and *no* universal definition exists. Here, we adopt a straightforward model to illustrate, in the next section, how energetic costs can arise in the operation of Maxwell's demon.

IV. PICKING A HOLE

Maxwell's demon was conceived by James Clerk Maxwell in 1871, as described in his book *Theory of Heat* [12]. This concept was introduced toward the end of the book, in a section discussing the limitations of the second law of thermodynamics. However, the idea originally appeared in a letter written by Maxwell in 1867, where he boldly proposed *"...picking a hole"* in the second law [13]. The famous term "demon" for this imaginary intelligent being was coined by Lord Kelvin 3 years after Maxwell's book. Although one might assume that this name was chosen to evoke a malicious creature, it was intended to highlight the idea of a supernatural being—a concept rooted in Greek mythology [79,80].

In this section, we introduce the ideas of Maxwell's demon and the Szilard engine. Rather than presenting the solution straight away, we take a more gradual, historical approach by first laying out the problem and leaving it open. Later, we bridge this discussion with tools from information theory, which allows us to resolve the apparent paradox. However, when discussing the Szilard engine, we also introduce new insights and results that provide a deeper understanding of this thought experiment.

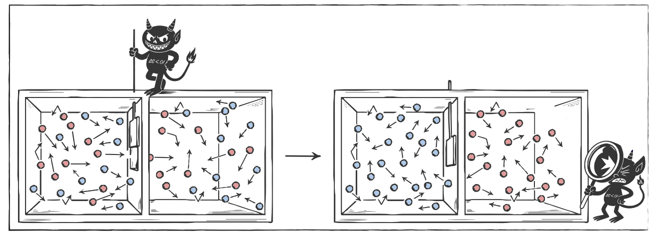


FIG. 2. *Temperature demon*. An intelligent being, a demon, controls a door between two gas chambers. As gas molecules approach the door, the demon selectively opens and closes it, allowing only fast molecules to pass one way and slow molecules the other. Since the kinetic temperature of a gas is directly linked to the velocities of its constituent molecules, the demon's strategy results in one chamber warming up while the other chamber cools down. This reduces the overall entropy of the system without requiring any work, thereby violating the second law of thermodynamics.

A. Maxwell's demon

We start by analyzing what is known as the *temperature demon*. The setup depicted in Fig. 2 illustrates the core concept of Maxwell's thought experiment: a tiny creature has access to a gas of particles that are initially in thermal equilibrium. The demon controls a trapdoor that it can freely open or close without friction. The creature's task is to generate a temperature difference in the gas without performing any work on it. Essentially, this scenario is analogous to producing a heat flow from a lower to a higher temperature without any other effect, which contradicts Clausius's formulation.

The kinetic theory of gases tells us that temperature is related to the average kinetic energy of the particles, and according to the Maxwell-Boltzmann distribution, some particles are faster or slower than the average [81]. Thus, the demon must be able to distinguish between the velocities of the particles that make up the gas. In other words, it is capable of measuring each particle, "remembering" what it measures, and then opening or closing the trapdoor accordingly. This implies that the *information* collected from the measurement is recorded in some sort of device, which we will refer to as *memory*.

If we assume that the demon itself is a physical system that is subject to the laws of thermodynamics, what do the first and second laws tell us about this setup? Since both the demon and the gas are thermally isolated, the demon's energy remains unchanged during the sorting process. According to the second law, however, the demon's entropy must increase by at least as much as the gas's

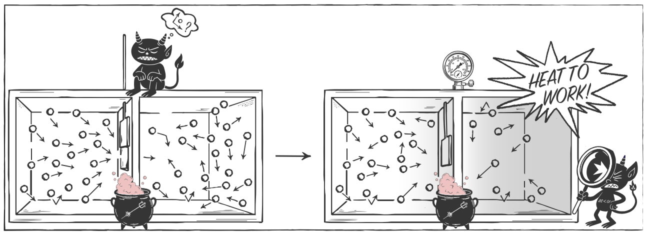


FIG. 3. *Pressure demon*. A less intelligent demon controls a door between two gas chambers connected by a constant temperature reservoir. As gas molecules approach the door, the demon selectively opens and closes it, allowing only right-moving molecules to pass in one direction and only left-moving molecules to pass in the opposite direction. This strategy increases the pressure on one side of the system while decreasing it on the other. The demon's actions convert heat from the reservoir into work, seemingly in violation of Kelvin's statement of the second law of thermodynamics.

entropy decreases. Therefore, the demon's entropy must increase while its energy remains constant.

Now, consider another demon that differs from the previous one by not being as smart—the *pressure demon* (see Fig. 3 for a pictorial representation). Instead of knowing the precise position and velocity of each particle, it only knows the direction in which each particle is moving. In this case, the demon allows all particles moving in one direction to pass through while stopping all those moving in the opposite direction. Effectively, this creates a pressure difference. It operates by making the gas interact with a heat bath at a constant temperature after generating a pressure inequality. Note that the sole effect of this process is the conversion of heat transferred from the heat bath into work, which constitutes a direct violation of Kelvin's statement of the second law. As we shall see in the next section, this type of demon was carefully discussed and analysed by Leo Szilard [14,15], who, for simplicity, considered a single particle instead of a gas.

Is it possible to design an autonomous machine that exploits statistical fluctuations to convert heat into work without a demon operating it? Smoluchowski [82,83] provided a negative answer to this question by demonstrating that a purely mechanical version of Maxwell's demon is impossible. A microscopic trapdoor designed to let only fast molecules pass would be affected by its own Brownian motion, causing it to move randomly. These thermal fluctuations would result in molecules passing between chambers in both directions, preventing any net energy transfer and making such a machine unreliable. Later, Feynman expanded on Smoluchowski's thought experiment to show that a Brownian ratchet, much like a microscopic trapdoor, would slip both forward and backward, preventing net work extraction [84,85].

Experimental Box 1: Maxwell's demon at work [86]

This experiment realizes Maxwell's demon while tracking all involved thermodynamic quantities, allowing full characterization of the demon's memory and its role in the thermodynamic process.

The demon is modeled as a microwave cavity, while the system is a superconducting qubit. They interact via the following Hamiltonian:

$$H = \hbar\omega_s \left|e\right\rangle\!\left\langle e\right| + \hbar\omega_D d^\dagger d - h\chi d^\dagger d \left|e\right\rangle\!\left\langle e\right| , \qquad (32)$$

where $|g\rangle$ and $|e\rangle$ represent the ground and excited states of the qubit, with an energy difference $h\omega_s$, d is the annihilation operator for a photon in the cavity, with resonant frequency ω_D and h is Planck's constant. From this Hamiltonian, we observe that no direct energy exchange occurs between the qubit and the cavity. However, the interaction induces a frequency shift of $-\chi$ in the cavity when the qubit is excited. If N photons are present in the cavity, this results in a total frequency shift of $-N\chi$.

The system and the demon's memory are initially prepared in thermal and vacuum states, respectively. A pulse at frequency ω_D is then applied to record the state of the system in the demon's memory. As described by the Hamiltonian, the cavity is excited only if the qubit is in the ground state. Work is extracted by applying a pulse at frequency ω_s, which acts on the qubit, functioning as a battery system. As indicated in Eq. (32), the demon controls the energy transfer by preventing the qubit from absorbing or emitting energy in an uncontrolled manner. If the correlation between the demon and the qubit is perfect, the energy flows from the qubit to the battery, ensuring that work is extracted. In this process, the system's entropy decreases while the demon's memory records the change. The demon's memory is then reset, completing the cycle.

A purely mechanical version of Maxwell's demon cannot work, but this does not rule out the possibility that such a violation occurs when an intelligent being, with information about the system, is present. In other words, what would happen to the second law of thermodynamics if we consider a mechanical demon equipped with a device, such as a computer, that gathers and processes information?

B. Szilard engine

The interplay between thermodynamics and information came up 62 years after Maxwell's demon was conceived. The Szilárd engine [14,15], a machine with a one-molecule working fluid, is a simple device that implements a cyclic process and extracts energy as work from a

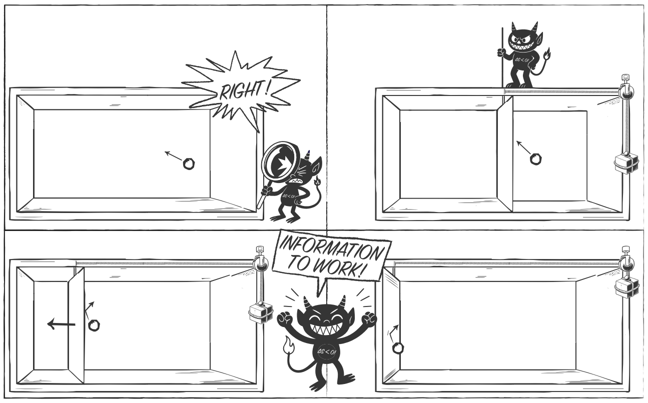


FIG. 4. *Szilárd engine*. A chamber containing a single particle (atom or molecule) is initially in an unknown position. A demon measures the particle's position and inserts a movable, massless wall with an attached mass. As the gas in the chamber expands at a constant temperature, the attached mass is raised, resulting in work being performed. This mechanism appears to enable the conversion of information into useful energy, seemingly contradicting the second law of thermodynamics, as the demon extracts heat from the heat bath and converts it fully into mechanical work.

thermal reservoir. As with Maxwell's demon, the Szilárd engine can apparently violate the second law of thermodynamics whenever some information about the state of the system is available.

Szilard's engine is a detailed version of Maxwell's pressure demon. Its work protocol is illustrated in Fig. 4. The engine consists of a chamber with volume V and a single gas molecule, in thermal equilibrium with a reservoir at temperature T, undergoing a three-step process. First, the demon measures the position of the particle, determining whether it is on the left or right side of the chamber. Second, based on the particle's position, it inserts a thin, massless, adiabatic partition into the chamber, dividing it into two halves. Importantly, a specific mass load is attached to the movable partition on the side where the molecule is located. Third, by maintaining the chamber at a constant temperature, the gas undergoes a quasistatic isothermal expansion, with the partition acting as a piston. When the partition reaches the end of the chamber, the gas returns to its original state, occupying the entire volume V. During this expansion, the heat extracted from the environment is completely converted into mechanical work.

The work extracted in the process can be calculated considering the following points: (i) the gas undergoes an isothermal expansion; (ii) the insertion and removal of the thin partition are performed reversibly, without expending energy; and (iii) the only energy contribution comes from the reversible expansion of the particle against the piston from $V/2$ to V. It should also be noted that the load must be continuously varied to match the gas pressure during expansion, ensuring that the process remains quasistatic and reversible. This condition allows us to express the pressure as $p = k_B T/V$. Thus, the amount of extracted work is given by

$$W = -\int_{V_i}^{V_f} p dV = k_B T \int_{\frac{V}{2}}^{V} \frac{dV}{V} = -k_B T \log 2. \qquad (33)$$

This simple protocol shows a direct conversion of heat into work, contradicts Kelvin's statement, and thus seemingly violates the second law of thermodynamics. However, it is worth noting that if the system is modified to include more partitions, the volume increase would be larger, leading to the potential for extracting more work. In this case, the work extracted would scale with the number of partitions, enhancing the total energy conversion.

Experimental Box 2: Single-electron box as a Szilard engine [87]

This experiment shows the extraction of $k_B T \log 2$ of work from a heat bath per bit of information, with a fidelity of 75%.

The setup consists of two metallic islands connected by a tunnel junction. These islands contain an electron gas, and an additional "extra" electron is introduced by adjusting the gate charge n_g. The difference in chemical potential between the islands is controlled by a gate voltage V_g, applied to one of them. Initially, V_g is set so that the extra electron is equally likely to be on either island. Next, the extra electron is detected using a single-electron transistor and is captured on one of the islands. To trap the electron, the gate charge is rapidly changed, resulting in an energy increase that prevents the electron from tunneling out. Finally, n_g slowly returns to its original condition, extracting energy from the heat bath in the process, and completing the cycle.

Unlike the classical Szilard engine, this realization involves a gas of electrons with an additional "extra" electron, instead of a single particle. After measuring the electron, no partition is installed. Instead, the charge configuration is changed, and the extracted work is not related to the volume but rather to the energy associated with changes in the gate charge.

If Szilard's conclusions are indeed correct, the system's entropy should decrease, making us wonder why that could not actually be the case. Interestingly, Szilard, in his writings, assumes that any measurement procedure is inherently associated with a certain amount of entropy production. When this entropy production is taken into account, it restores the validity with the second law. Moreover, he argues that the amount of entropy produced could be greater than, but not less than, the value given by Eq. (33). This is a remarkable observation, as we will see in Sec. VI, where we shall connect it with a fundamental bound on computation.

Although not entirely convinced, Szilard suspected that the act of measuring and recording information in a memory could eventually explain such a possible violation. Motivated by this, he considered two additional models that involve memory. However, it remained unclear whether the thermodynamic cost arose from the measurement, the act of recording, or the process of forgetting. Nevertheless, memory was regarded to play a crucial role in the demon's operation. Importantly, Szilard's reinterpretation of Maxwell's demon contributed to the concept of a bit of information. Years later, this notion would become a central concept in information and computing theories.

However, we may question the conclusions of Szilárd's engine, given that the entire analysis relies on a single molecule, where fluctuations cannot be neglected, and traditional thermodynamic methods may not apply. This objection can be overruled by noting that linking multiple Szilárd engines together minimizes fluctuations, making it possible to apply thermodynamics. Yet, even if we accept the single-molecule system, the gas behaves in a way that seems to violate the ideal gas law. When the partition is inserted into the chamber, the gas is suddenly confined to half of its original volume without any change in temperature or expenditure of energy.

The second objection is more difficult to refute than the first. Szilárd's position holds that, although the molecule is on one side of the partition after it is inserted, the key point is that we do not know which side it is on until a measurement is made. This uncertainty is critical because, from an informational perspective, the gas as a system still effectively occupies both sides of the partition. However, one might question whether this argument is subjective, as it depends on the observer's information (see Ref. [88] for an extended discussion). Until we measure which side of the partition the molecule is on, we cannot fully define the system's state or extract any useful work. This blurs the line between an objective thermodynamic process—one that should operate independently of the observer—and a subjective one, where the observer's knowledge plays a role. The apparent subjectivity is resolved by appealing to stochastic thermodynamics [26,89]. Before any measurement, the molecule is described by an objective probability distribution over the two half-boxes, and the associated free energy already accounts for the cost of inserting the partition. A fully quantum treatment by Zurek [90,91] reaches the same conclusion and, by invoking spatial superposition and decoherence, makes the observer independence of the argument explicit at the microscopic level.

A third objection concerns the assumption that the insertion and removal of the partition do not incur any energy cost [92]. While this assumption is valid in the classical case, it does not hold when the problem is approached quantum mechanically. This is evident if one considers Szilard's setup as a single particle of mass m confined within a one-dimensional square-well potential of length L. Solving the time-independent Schrödinger equation yields the following energy spectrum for the particle $\epsilon_n = \hbar^2\pi^2 n^2/2mL^2$. Consequently, when the partition is introduced—a process that can be modeled as a potential barrier erected in the middle of the box—the boundary conditions change. As a result, the energy spectrum is altered to $E'_n = \hbar^2\pi^2 n^2/2m(L/2)^2$. Using the decomposition provided by Eq. (29), one observes that $d\epsilon_n$ is nonzero due to the shifting energy levels. This implies that dW is also nonzero, according to Eq. (31), which quantifies the work performed during this process. The next question is whether this process can be assumed to be adiabatic. Using again Eq. (29), this would imply that $dp_n = 0$, meaning that the occupation probabilities remain constant during the process. However, as discussed, the energy levels shift nonuniformly when the partition is inserted. Consequently, the ratio p_n/p_m no longer satisfies the Boltzmann distribution, indicating that the system is no longer in thermal equilibrium. However, since the system interacts with a heat bath, thermalization will occur, allowing the entire process to be described as adiabatic and isothermal.

✎ *Quantum Szilard engine*

Consider N identical particles prepared in a potential well of size L (see Fig. 5) and the four-step protocol: (i) partition insertion, (ii) measurement, (iii) expansion, and (iv) partition removal. Unlike the classical Szilárd engine, we insert the partition before making any measurements. This is justified because, before measurement, the particles are delocalized across the entire box, with a probability amplitude describing their likelihood of being found at any position. Following Ref. [92], we calculate the thermodynamic cost associated with each step.

(i) A wall is isothermally inserted in a position l. At this moment, the partition function $Z(l)$ includes all accessible microstates. Since the system is in a superposition of states with different numbers of particles m on the left and $N-m$ on the right, the system's partition function is $Z(l) = \sum_{m=0}^{N} Z_m(l)$. The work required for the insertion process is then

$$W_{\rm ins} = \frac{1}{\beta}\left[\ln Z(l) - \ln Z(L)\right]. \qquad (34)$$

(ii) We perform a measurement, collapsing the superposition and projecting the system into a specific state with a definite number of particles m on the left and $N-m$ on the right. The measurement is assumed to be perfect and does not require any

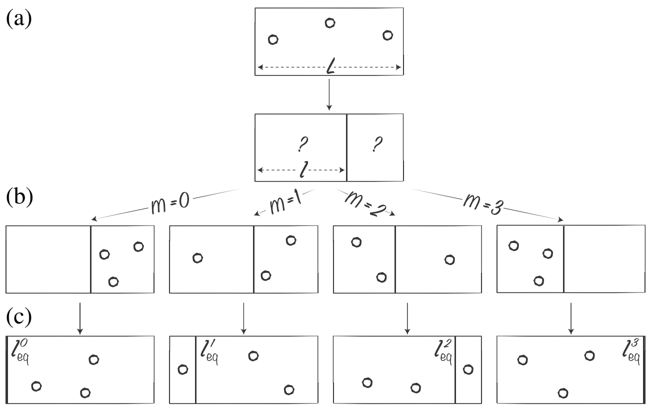


FIG. 5. *Quantum Szilárd engine*. Three quantum particles are confined in a box of size L. (a) A partition is installed at location l, and (b) a measurement is made revealing the number m of particles on the left. (c) Then, the particles expand isothermally until they reach equilibrium at l^m_{eq}, and the partition is removed.

work. Consequently, the system can be in any of the possible states characterized by different m with probability $p_m = Z_m(l)/Z(l)$.

(iii) The system undergoes an isothermal expansion. The wall moves until it reaches an equilibrium position l^m_{eq} determined by the balance of forces, $F^{\text{left}} + F^{\text{right}} = 0$, where the generalized force F is defined as $\sum_n P_n(\partial\epsilon_n/\partial X)$. The average work extracted during the expansion is calculated by finding the work for each possible m weight it by its probability p_m, namely

$$W_{\text{exp}} = \frac{1}{\beta}\sum_{m=0}^{N} p_m \left[\ln Z_m(l^m_{\text{eq}}) - \ln Z_m(l)\right]. \tag{35}$$

(iv) The wall is removed isothermally and quasistatically [93]. The average work in the removal process is

$$W_{\text{rem}} = \frac{1}{\beta}\sum_{m=0}^{N} p_m \left[\ln Z(L) - \ln Z(l^m_{\text{eq}})\right]. \tag{36}$$

Observe that the second term of Eq. (36) accounts for the partition function $Z(l^m_{\text{eq}}) = \sum_{m=0}^{N} Z_m(l_{\text{eq}})$ as after removal, all particles are once again delocalized [compare this with Eq. (35)].

Combining all the contributions, the total work during the cycle is given by

$$W_{\text{tot}} = W_{\text{ins}} + W_{\text{exp}} + W_{\text{rem}} = -\frac{1}{\beta}\sum_{m=0}^{N} p_m \ln\left(\frac{p_m}{p^*_m}\right), \tag{37}$$

where $p^*_m = Z_m(l^m_{\text{eq}})/Z(l^m_{\text{eq}})$. Let us analyze Eq. (37). First, note that the classical scenario is recovered by setting $N = 1, l = L/2$. This implies that $p_0 = p_1 = 1/2$ and since $m = 0, 1$, the wall reaches the end of the box so that $Z(l^m_{\text{eq}}) = Z_m(l^m_{\text{eq}})$ and $p^*_m = 1$ are always true. Consequently, Eq. (37) reads $W_{\text{tot}} = \beta^{-1}\log 2$. We can also compute each term independently to show that $W_{\text{ins}} = \beta^{-1}\ln 2 - \Delta$, where $\Delta = \log[z(L)/z(L/2)]$ with $z(l) = \sum_{n=1}^{\infty} e^{-\beta\epsilon_n(l)}$ and $\epsilon_n(l) = \hbar^2\pi^2 n^2/(2ml^2)$; and similarly, $W_{\text{exp}} = \Delta$ and $W_{\text{rem}} = 0$. As temperature increases, the classical results of individual steps are recovered, i.e., $W_{\text{ins}} \to 0$, $W_{\text{exp}} \to k_B T\log 2$, and $W_{\text{rem}} = 0$, since Δ approaches $k_B T\log 2$ in this limit.

Another interesting case to analyze is when $N = 2$ and $l = L/2$. In this setup, $p^*_0 = p^*_2 = 1$, and for $m = 1$, the wall does not move during the expansion process, as $l = l^1_{\text{eq}}$ and $f_1 = p^*_1$. This leads to $W_{\text{tot}} = -2\beta^{-1}p_0\log p_0$. Ignoring particle spin and considering the extreme cases of temperature, we find that for $T \to 0$: $W_{\text{tot}} = 2/3\log 3$ for bosons, as they can occupy the same state, and $W_{\text{tot}} = 0$ for fermions, due to the Pauli exclusion principle. The other limit, $T \to \infty$, both bosons and fermions yield $W_{\text{tot}} = \log 2$, as the particles behave classically at high temperatures.

✎

The quantum Szilard engine, as described in the example above, sparked considerable discussion regarding the role of particle statistics and the relationship between information gain and extractable work. The assumptions and conclusions of Ref. [92] were later critiqued in Ref. [94], which questioned the handling of the partition removal stage and suggested that the original protocol could lose potential work due to tunneling effects. The critique proposed an alternative approach, emphasizing the role of generalized forces during partition removal and suggesting that the extracted work should always be non-negative. In their reply, the original authors defended their approach, arguing that the critique misunderstood the role of partition function discontinuities and incorrectly characterized their protocol as nonoptimal [95]. They clarified that under their protocol, optimal conditions ensure positive work extraction, and any negative results stem from misinterpreting the nonoptimal scenarios.

The discussion culminated in a subsequent work [96], which reframed the debate by connecting the extraction of work with the gain in information during the measurement phase. This work showed the following: (i) The total extractable work, is determined solely by the mutual information gained from the measurement, regardless of particle type or statistics. (ii) Quantum effects influence intermediate stages (e.g., expansion and removal), but when averaged, total work depends only on the number of measurement outcomes.

C. Efforts to patch the hole

The measurement aspect raised by Szilard was further explored in the 1950s by Leon Brillouin [20] and Dennis Gabor [97]. Independently, both demonstrated that the information obtained by performing a measurement results in an increase in entropy. To understand the main argument behind this claim, imagine that in Szilard's engine, we want to determine the position of the gas particle. The chamber containing this single gas particle is at a constant temperature, and, optically speaking, the radiation inside is that of a black body. According to quantum theory, this electromagnetic radiation, composed of photons, has a well-defined energy distribution [98]. Hence, a high-temperature lamp is sufficient to provide light signals (photons) distinguishable from the existing blackbody radiation, enabling the detection of the gas particle's position.

Brillouin then concluded that introducing a light source to observe the particle is accompanied by an increase in entropy, which is sufficient to save the second law of thermodynamics. More precisely, the measurement process would dissipate energy on the order of $k_B T$. A bold comment we might make here is that Brillouin's reasoning is based on a vicious circle: if we assume that the chamber in which the demon operates is a blackbody, then it follows Kirchoff's law [99]—however, this law is derived using the second law of thermodynamics. Consequently, it is no surprise that we can prove the second law to be satisfied.

Interestingly, Brillouin tackled this problem mathematically and, inspired by Shannon's recent theory of information (discussed in the next section), treated information and thermodynamics on an equal footing. This fusion provided a mathematical basis for proving the increase in entropy. It is amusing to compare this to Maxwell's original idea—Brillouin actually narrowed the demon's capabilities, as the demon now requires a physical means to obtain information about the particle's position. Ironically, in their attempt to resolve Maxwell's puzzle, Brillouin and Gabor discarded the role played by the demon's memory.

V. INFORMATION AND COMPUTATION INTERLUDE

Szilard's engine planted the seeds of modern information theory two decades before the formal mathematical framework emerged. This setup can be seen as a device that acquires information and trades it for work extraction, seemingly in contradiction to the second law. Brillouin, however, mathematically argued that information gathering is inherently linked to an increase in entropy. What is more, he treated *informational entropy* and thermodynamic entropy as fundamentally equivalent. But what exactly is informational entropy, and how does it relate to thermodynamics?

From another angle, we can break down Szilard's engine into a simple three-step process: the demon gets some information, makes a decision based on the acquired information, and then performs a protocol. So, could we loosely argue that Szilard's engine is a thermodynamic model of computation? After all, it is essentially processing information (where the molecule is), making a logical decision (which side to open), and carrying out an action (allowing the gas to expand and do work). The whole process feels very much like a computation happening through a physical system.

This section builds on these points to strengthen the connection between Szilard's engine, information theory, and computation. We begin with a recap of Claude Shannon's mathematical theory of information [17,18], followed by brief remarks on the concept and evolution of computation.

A. Classical information processing

When we think about information, it is easy to imagine an endless stream of zeros and ones, just like in The Matrix [100]. But beyond the cool visuals, this collection of binary digits obeys its own set of rules, known as Boolean algebra [101]. Developed in the mid-19th century, this toolkit provided a new theoretical playground for logic and philosophy. Most importantly, operations such as AND, OR, and NOT make it possible to manipulate bits of data in surprisingly simple ways (see Fig. 6). However, applying these concepts to real-world processes was still a distant reality. It took 84 years for Boolean algebra to find its practical application in the design of digital circuits, thanks to Shannon's groundbreaking master's thesis [17]. Shannon bridged the gap between Boolean logic and electrical circuits by demonstrating that the rules of Boolean algebra could be used to design switching circuits. By treating the ON and OFF states in electrical circuits as 1 (true) and 0 (false), he showed that Boolean logic operations could be directly implemented using relays and switches. In other words, computation, once primarily understood as the manipulation of numbers and symbols to solve mathematical problems or to reason logically, is now seen as more physically and practically—linked to the manipulation of binary information in electrical circuits. This shift made computation synonymous with the automatic and efficient processing of data by digital computers. With this in mind, one might now ask a deeper question: what exactly is information?

To illustrate the modern concept of information, consider two different coins: one always lands heads, while the other lands randomly, sometimes heads, sometimes tails. Before flipping the first coin, we already know the result, so we learn nothing from flipping it. In contrast, before flipping the second coin, our uncertainty is at its maximum, and by observing the result, we gain new information. This example shows that information

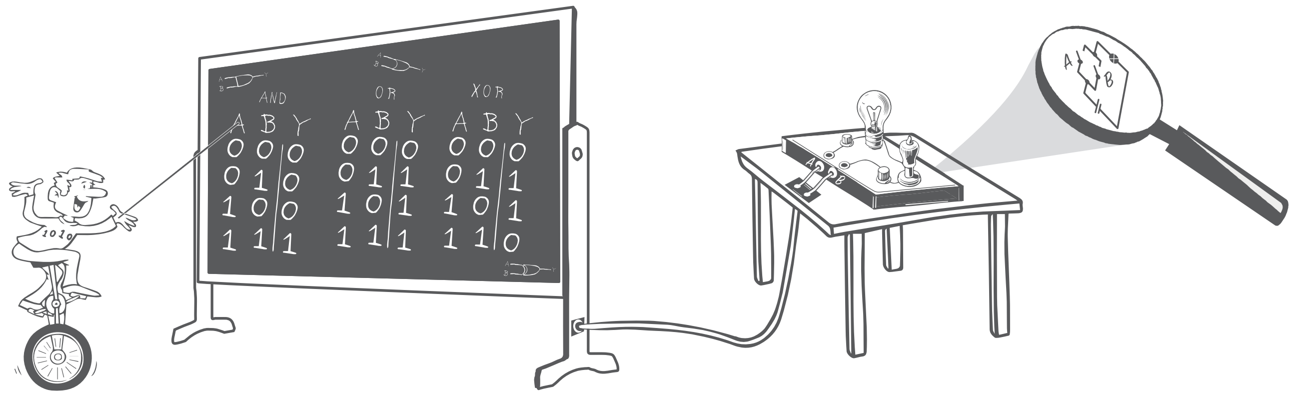


FIG. 6. *A symbolic analysis of relay and switching circuit*. Shannon, famously known for riding a unicycle around the halls of Bell Labs, demonstrated that Boolean algebra could be implemented using relay and switching circuits. Boolean algebra consists of three elementary operations: AND, OR, and NOT. The rules for these operations are outlined in the truth table sketched in Shannon's blackboard. For example, the implementation of an OR gate can be represented as a simple electrical circuit. In this case, switches A and B are arranged in parallel, meaning the circuit will be completed (and the light bulb will turn on) if either switch A, switch B, or both are closed.

can be understood as both the uncertainty we have before observing an event and the knowledge we gain afterwards. Information relates to uncertainty, and since uncertainty is linked to disorder, this establishes an initial connection between information and entropy, albeit in a rudimentary form. To formalize this intuition about information, in what follows we will adopt an axiomatic approach similar to that of Shannon in his seminal 1948 article [18], which introduced a mathematical theory of communication.

If a given event has a probability p, we expect that the amount of information contained in such an event should satisfy certain properties. First, it should be label independent, meaning that the information content of an event is determined entirely by how likely or unlikely the event is, regardless of what the event is called or how it is labeled, as long as the mapping between the labels is a bijection. Thus, it depends only on the probability p of the event. Second, the amount of information $i(p)$ for a given event should be greater the less likely the event is to occur—less likely an event is to occur, the more information it provides when it happens. Mathematically, the amount of information $i(p)$ should be a continuous and monotonically decreasing function of p as the probability of an event increases (making the event more likely), the information content decreases. This reflects the idea that more likely events are less surprising and, therefore, carry less information. Importantly, in the extreme cases where an event is completely certain (with probability $p = 1$), the amount of information is zero because no surprise occurs when the event happens, and in the case of an impossible event (with probability $p = 0$), the amount of information should be infinite, as you would be infinitely surprised to observe such an event. This can also be understood through a continuity argument, where as p approaches zero, the information content grows without bound. Third, the information from two independent events with probabilities p and q should sum up to $i(p) + i(q)$. This additivity captures the idea that when two independent events occur, the total information gained is simply the sum of the information from each event individually. Since the probability of two independent events occurring is the product of their individual probabilities, pq, the additivity condition implies that the information from such events should satisfy $i(pq) = i(p) + i(q)$. The only function that satisfies this property is the logarithm. This implies that, up to a constant factor k, the information content of an event with probability p should be given by

$$i(p) = -k \log p. \tag{38}$$

The constant factor k can be understood as a scaling factor that adjusts the units in which we measure information, often chosen to make the mathematical expressions more convenient, such as when using logarithms with base 2, 10 or e. Observe that the function in Eq. (38) is the only one satisfying the three properties [102]. The logarithm depends solely on the probability p and is independent of any specific labels or characteristics of the event. It decreases as p increases, and the factor $-k$ ensures that $i(p)$ remains positive and decreasing. Finally, the logarithmic function naturally exhibits additivity, which means that the information content from two independent events sums appropriately, making it the ideal candidate for quantifying information.

Consider a random variable X described by a probability distribution $p(X = x) := p_X(x)$. The information content of a specific outcome x, given by $-k \log p_X(x)$, measures how much information is gained when that outcome occurs. To capture the overall uncertainty or the average information content associated with the entire distribution of X, we sum up this quantity across all possible outcomes, weighted by their probabilities. Setting $k = 1$ in the expression for information content, we obtain the famous Shannon entropy (or information entropy):

$$S(X) := -\sum_x p_X(x) \log p_X(x). \tag{39}$$

We slightly abuse the notation by representing thermodynamic, Shannon, and von Neumann entropies with the same letter. While this choice is justified in the case of Shannon and von Neumann entropies, as any probability distribution naturally corresponds to a diagonal quantum state, we also use it for the Gibbs entropy because, in the scenarios of interest, these notions coincide (see discussion below).

Even though they come from different contexts, the Shannon and von Neumann entropies share a striking similarity—they both measure uncertainty or "missing information." In fact, von Neumann entropy can actually be thought of as a natural extension of Shannon entropy. This becomes clear if we imagine a source that prepares messages composed of n letters, each letter chosen from an ensemble of quantum states represented by ρ. The probability of any measurement outcome for a letter from this ensemble—assuming the observer has no knowledge of the specific letter prepared—can be fully characterized by the density operator ρ. When we select an orthonormal basis that diagonalizes ρ, the vector of eigenvalues forms a probability distribution, and the von Neumann entropy then corresponds to the Shannon entropy of this distribution. However, it is important to emphasize that von Neumann entropy serves multiple roles. It not only quantifies the quantum (and classical) information content per letter of a pure state ensemble but also measures the entanglement of a bipartite pure state, among various other applications in quantum information theory (for a detailed discussion, see the books by Bengtsson and Zyczkowski [103], Wilde [104], as well as the lecture notes by Preskill [105]).

Interestingly, as mentioned in Sec. III, von Neumann entropy was introduced almost 20 years before Shannon's. The funny thing is that Shannon was not sure what to call his function at first. He thought about calling it by either "information" or "uncertainty", but both terms were already overused in the literature. That is when von Neumann stepped in and gave him this advice [106]:

"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, no one really knows what entropy really is, so in a debate you will always have the advantage".

While this anecdote is often cited, its historical accuracy is anecdotal rather than rigorously documented. More importantly, the mathematical and conceptual connection between information-theoretic entropy and statistical mechanics (e.g., Gibbs entropy) plays a meaningful role in modern thermodynamics—a connection we discuss at the end of this section.

For continuous random variables, the sum in Eq. (39) is replaced by an integral. It follows directly from Eq. (39) that for an identically distributed binary variable, where $p(x=0) = p(x=1) = 1/2$, $S(X) = 1$ (using a base-2 logarithm), representing the unit of information known as a bit. Note that Eq. (39) bears a mathematical resemblance to the entropy in the canonical ensemble, $S = k_B \log \Omega$, where Ω represents the number of possible microscopic arrangements of the atoms or molecules in a thermodynamic system. However, these two expressions are not merely similar; they coincide (up to a constant factor) within the assumptions of the microcanonical ensemble. Specifically, when there are Ω equiprobable microstates, the corresponding probability is $p_x = 1/\Omega$, leading to the same form as Eq. (39). The derivation and interpretation of these two types of entropy arise from very different contexts. Recall from the previous section that Brillouin boldly postulated a direct connection between the two [107].

Reasonably, one can define a whole zoo of entropies based on the above formulation. For example, when we discuss information—particularly in the context of extracting it from a system—we are usually concerned with determining the state X of the system by measuring a related quantity Y. The measurement of Y yields information about X, thus reducing the uncertainty associated with X. The degree to which uncertainty is reduced represents the amount of information that Y provides about X. Mathematically, this can be expressed in terms of the mutual information:

$$I(X:Y) := S(X) - S(X|Y), \tag{40}$$

where $S(X|Y)$ represents the uncertainty of the posterior probability distribution $p(x,y)$ averaged over the possible outcomes, defined as

$$\begin{aligned} S(X|Y) &:= \sum_y p_Y(y) \left[-\sum_x p_{X|Y}(x|y) \log p_{X|Y}(x|y) \right] \\ &= -\sum_{x,y} p_{XY}(x,y) \log p_{X|Y}(x|y). \end{aligned} \tag{41}$$

Observe that we previously defined mutual information [Eq. (40)] using the von Neumann entropy in Sec. III when discussing correlations between two subsystems. Using

Bayes' formula,

$$p_{XY}(x,y) = p_{X|Y}(x|y)p_Y(y), \tag{42}$$

the mutual information can be rewritten as

$$I(X:Y) := -\sum_{x,y} p_{XY}(x,y)\log\left[\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}\right]. \tag{43}$$

From this expression several important properties of mutual information can be derived. First, it is symmetric, meaning that the amount of information X provides about Y is identical to what Y provides about X. Second, by applying the properties of logarithms, it can be shown from Eq. (43) that $I(X)$ is always non-negative and becomes zero only when X and Y are statistically independent. This makes mutual information a reliable measure of the correlation between X and Y. Third, if we measure a quantity $Y = f(X)$ perfectly, without errors, then $S(X|Y) = 0$ and $I(X) = S(Y)$. In other words, the information gained from an error-free measurement equals the uncertainty of the measured outcome. Lastly, Eq. (43) offers three different ways to represent mutual information:

$$\begin{aligned} I(X:Y) &= S(X) - S(X|Y) && (44)\\ &= S(Y) - S(Y|X) && (45)\\ &= S(X) + S(Y) - S(X,Y) \geq 0, && (46)\end{aligned}$$

where $S(X,Y) := -\sum_{x,y} p_{XY}(x,y)\log p_{XY}(x,y)$ is the joint Shannon entropy. The last equality shows that correlations cause the entropy to be subadditive, meaning $S(X,Y) = S(X) + S(Y) - I(X:Y)$. These relationships can be easily visualized using a Venn diagram, as shown in Shannon's blackboard in Fig. 7. This expression will be useful in Sec. VII when we explore the physical nature of Maxwell or Szilárd demons and interpret a measurement as the creation of correlations between the state of the demon and the state of the system.

As a final question in this section, one might ask whether there is a connection between Shannon, von Neumann, and thermodynamic entropy. The third example discussed in Sec. II shows that the von Neumann entropy (or Shannon entropy, when the density operator is expressed in the energy eigenbasis) coincides with the thermodynamic entropy when the probability distribution corresponds to the canonical ensemble. Note that this conclusion holds only if the system satisfies the equivalence of the ensembles [108]. What about for general states? Attributing entropy to a nonequilibrium state has, in fact, been a prominent topic of research for many years [109]. In general, identifying thermodynamic entropy with von Neumann entropy is incorrect—a point originally noted by von Neumann himself [110].

As shown in Sec. III, a pure energy eigenstate of an isolated system has zero von Neumann entropy,

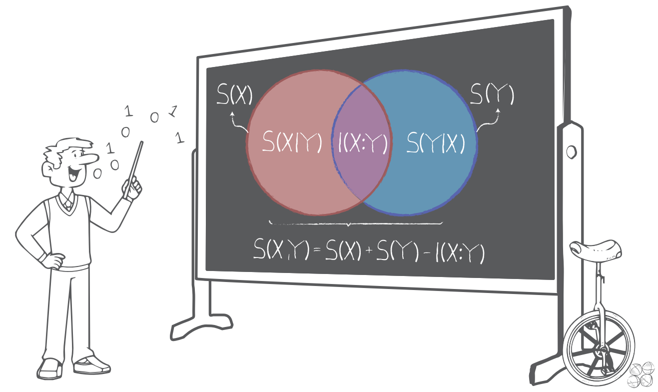


FIG. 7. *Information diagram*. Venn diagram illustrating the relationships among the joint entropy $S(X,Y)$, individual entropies $S(X)$ and $S(Y)$, conditional entropies $S(X|Y)$ and $S(Y|X)$, and mutual information $I(X:Y)$. Next to it, we see Shannon's unicycle and some of his juggling balls, as he was known for often doing both simultaneously.

whereas its equilibrium description—via a microcanonical or canonical ensemble—generally has nonzero thermodynamic entropy. A qualitatively different example arises in strongly coupled or non-Markovian setups, where the system and environment cannot be cleanly separated into distinct energy contributions. In such regimes, even if the global state has well-defined thermodynamic properties, the von Neumann entropy of the subsystem may not coincide with its thermodynamic entropy [75,76]. Further discrepancies appear in bipartite entangled systems: although the global state may be pure (with zero von Neumann entropy), each subsystem can exhibit nonzero von Neumann entropy. In such cases, the thermodynamic interpretation is ambiguous without an explicit equilibrium framework. These diverse scenarios highlight that, outside of the weak-coupling, equilibrium regime, von Neumann entropy generally does not coincide with thermodynamic entropy.

However, in both stochastic [26,111,112] and quantum thermodynamics [28,30,113] the von Neumann (and Shannon) entropy has shown a clear physical meaning in specific contexts: it governs the energetics of nonequilibrium processes for systems weakly coupled to one or more thermodynamic reservoirs. If we now wonder whether there is an entropic function capable of describing the thermodynamic entropy, it has been discussed that the observational entropy is the most appropriate candidate as it unifies the Gibbs-Shannon-von Neumann entropy and the Boltzmann one [114–118]. Observational entropy is defined with respect to a coarse graining of the system's Hilbert space into a set of orthogonal projectors $\mathcal{X} = \mathbb{P}_x$. Given a quantum state ρ, it is expressed as

$$S_{\text{obs}}^{\mathcal{X}} = \sum_x p_x(-\log p_x + \mathcal{V}_x), \tag{47}$$

where $p_x = \mathrm{tr}(\mathbb{P}_x\rho)$ is the probability of obtaining outcome x, and $\mathcal{V}_x = \mathrm{tr}(\mathbb{P}_x)$ quantifies the volume of the subspace associated with $\mathbb{P}_x$. This expression captures both how the state is distributed across different macrostates and how large each macrostate is. By choosing the coarse graining appropriately, observational entropy can reproduce the von Neumann entropy (when the coarse graining is very fine) or the Boltzmann entropy (when the state is localized in a single macrostate). Unlike the von Neumann entropy, it can increase under unitary evolution when the coarse-graining is fixed, making it especially suitable for describing entropy production in isolated systems. For this reason, it has been proposed as a generalized definition of thermodynamic entropy, applicable even far from equilibrium. For a detailed discussion of observational entropy, we suggest the tutorial by Strasberg and Winter [109], which provides an accessible introduction and reviews how various entropy definitions behave in quantum thermodynamics. We also point to Ref. [119], which very recently introduced a general framework unifying observational entropy with maximum entropy principles, and showed how it captures entropy production across a broad range of scenarios.

B. Thermodynamics of information

Let us now combine elements of information theory, such as the notions of Shannon and von Neumann entropy, with thermodynamic concepts. Our main motivation for doing so comes from the fact that Maxwell's work revealed an interplay between thermodynamic entropy and information. However, traditional formulations of the second law, such as those of Clausius and Kelvin, do not address information. Here, we first provide elements for explicitly incorporating information into thermodynamics and then investigate the possible thermodynamic costs associated with manipulating information, including processes such as measurement, erasure, copying, and feedback. The following discussion is focused on our goal of resolving Maxwell's demon. For a broader discussion of the thermodynamics of information, we recommend the following readings [59,120–123].

In Sec. III, we extended the concept of free energy beyond equilibrium states [Eq. (25)]. Specifically, for a system described by a Hamiltonian H and prepared in a state ρ in the presence of a heat bath at an inverse temperature β, this quantity is defined as [124]

$$\mathcal{F}(\rho, H) := \langle H\rangle_\rho - \frac{1}{\beta}S(\rho), \tag{48}$$

where the first term $\langle H\rangle_\rho$ denotes the average energy and $S(\rho)$ the Shannon or von Neumann entropy, depending on the context. For classical systems, where x labels the microstate and $\rho(x,t)$ is the probability density over phase space, the average energy becomes $\langle H\rangle_\rho = \int \mathrm{d}x\,\rho(x,t)H(x)$ and S corresponds to the Shannon entropy. In contrast, for quantum systems, ρ is a density matrix and the expressions reduce to $\langle H\rangle_\rho = \mathrm{tr}(\rho H)$ and $S(\rho)$ becomes the von Neumann entropy.

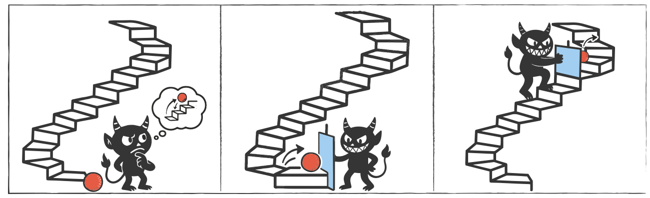


FIG. 8. *Information-driven motion*. A microscopic particle moves within a spiral-staircase potential, where thermal fluctuations cause it to randomly jump between steps. By measuring the particle's position at regular intervals, we can determine when it jumps upwards. This allows us to implement a feedback mechanism by placing a block behind the particle, preventing subsequent downward jumps. Repeating this process allows the particle to move up the staircase. Ideally, the energy required to position the block is negligible, meaning the particle's motion is driven purely by the information obtained from measuring its position. Figure inspired by Ref. [125].

As we previously discussed, the nonequilibrium free energy provides a bound for the minimum average work required to isothermally drive the system from one arbitrary state to another (or the maximum average work that can be extracted from the system when it is in an out-of-equilibrium state):

$$W \geq \Delta\mathcal{F}. \tag{49}$$

To give a gist of what is to come, we consider a motivating example, which was experimentally realized in Ref [125]: a microscopic particle on a potential shaped like a spiral staircase (see Fig. 8). The height of each step is comparable to k_BT. The particle, driven by thermal fluctuations, jumps between steps—sometimes moving up, sometimes down. Now, consider the following feedback control: the particle's position is measured, and if an upward jump is observed, a thin partition is placed behind it to prevent a downward jump. This partition can be installed at negligible energy cost. By repeating this control at every jump, the particle is expected to keep climbing up the stairs.

Again, at first glance, it seems as though the particle is gaining free energy from nowhere, apparently violating the second law of thermodynamics, much like Maxwell's demon or Smoluchowski and Feynman's ratchet. However, the particle's movement is actually driven by the information obtained from measuring its position. Here, the feedback control rectifies thermal fluctuations. The energy gained from the information is balanced by the energy cost to the demon for manipulating this information. When we consider the total system, including both the particle and the demon, the second law holds. In this scenario, the demon simply consists of macroscopic devices, like our computers.

We now introduce the main tools for modeling thermodynamic processes when information is involved.

✎ *Feedback process*

Consider a protocol in which a system, interacting with a heat bath at inverse temperature β, is measured, and the obtained information is then used to perform an isothermal process. To determine the thermodynamic cost of acquiring this information, we assume a classical system with a continuous random variable x representing its microstate. Initially, the system is described by a Hamiltonian $H(x,t)$ and a probability density $\rho_X(x,t)$. At the very start, before any measurement or external interaction, the system's free energy is $\mathcal{F}_{\text{in}}$. At time $t=\tau$, an agent measures a quantity M and gets a discrete outcome m. Just right before the measurement, its free energy, denoted by $\mathcal{F}_{\text{pre}} := \mathcal{F}[\rho_X(x,\tau), H(x,\tau)]$, is given by

$$\mathcal{F}_{\text{pre}} = \langle H(x,\tau)\rangle_{\rho_X} - \frac{1}{\beta}S(X), \tag{50}$$

where $\langle H(x,\tau)\rangle_{\rho_X} = \int \mathrm{d}x H(X)\rho_X(x,\tau)$. Typically, we distinguish between the initial and premeasurement free energies, as the system might not remain static; it could undergo controlled or natural dynamics leading up to the measurement point, thereby changing its state and, consequently, its free energy. The system's state after the measurement is then updated using Bayes' rule, resulting in

$$\rho_{X|M}(x|m) = \frac{p_{M|X}(m|x)\rho_X(x,\tau)}{p_M(m)}. \tag{51}$$

Since we are in a classical setting, the measurement does not disturb the system. This is captured by demanding that the system's state remains unchanged after the measurement

$$\sum_m p_M(x)\rho_{X|M}(x|m) = \rho_X(x,\tau). \tag{52}$$

Using Eq. (48), we can write the free energy of the system after the measurement for a given outcome m as

$$\mathcal{F}[\rho_{X|M}(x|m), H(x,\tau)] = \langle H(x,\tau)\rangle_{\rho_{X|M}} - \frac{1}{\beta}S[\rho_{X|M}(x|m)], \tag{53}$$

where $\langle H(x,\tau)\rangle_{\rho_{X|M}} = \int \mathrm{d}x H(x,\tau)\rho_{X|M}(x|m)$. If we take the average over all possible outcomes, we obtain the nonequilibrium free energy of the system after the measurement

$$\begin{aligned}\mathcal{F}_{\text{pos}} &:= \sum_m p_M(m)\mathcal{F}[\rho_{X|M}(x|m), H(x,\tau)] \\ &= \int \mathrm{d}x\, H(x,\tau)\rho(x,\tau) - \sum_p p_M(m)S[\rho_{X|M}(x|m)] \\ &\overset{(41)}{=} \langle H(x,\tau)\rangle_{\rho_X} - \frac{1}{\beta}S(X|M).\end{aligned} \tag{54}$$

Thus, the free-energy difference between after and before the measurement is determined using Eqs. (54)–(50):

$$\begin{aligned}\Delta\mathcal{F}_{\text{meas}} :&= \mathcal{F}_{\text{pos}} - \mathcal{F}_{\text{pre}} = \frac{1}{\beta}[S(X) - S(X|M)] \\ &= TI(X:M).\end{aligned} \tag{55}$$

Since mutual information is non-negative, measurement (or information acquisition) always increases the free energy, thus raising the amount of work that can be extracted isothermally. After the measurement, the agent gains information and uses it to perform the subsequent steps, a process we refer to as a *feedback process*. Consequently, we define the system's final free energy, $\mathcal{F}_{\text{fin}}$, after the feedback control has been applied based on the measurement outcome. Similarly to before, $\mathcal{F}_{\text{fin}}$ may differ from $\mathcal{F}_{\text{post}}$ because the feedback operation can either perform work on the system or extract work from it, depending on the protocol.

Finally, we can bound the work associated with the feedback process using the free energy bound given by Eq. (49). The total work done throughout the entire feedback process includes both the stages before and after the measurement. Consequently, we can break down the process into two subprocesses: from the initial to the premeasurement state, $\mathcal{F}_{\text{init}} \to \mathcal{F}_{\text{pre}}$, and from the postmeasurement to the final state, $\mathcal{F}_{\text{post}} \to \mathcal{F}_{\text{fin}}$. By applying Eq. (49) to each subprocess, we obtain

$$W_{\text{fb}} \geq (\mathcal{F}_{\text{fin}} - \mathcal{F}_{\text{post}}) + (\mathcal{F}_{\text{pre}} - \mathcal{F}_{\text{init}}) = \Delta\mathcal{F} - \Delta\mathcal{F}_{\text{meas}}, \tag{56}$$

with $\Delta\mathcal{F} := \mathcal{F}_{\text{fin}} - \mathcal{F}_{\text{init}}$ being the difference between the final and initial free energy states immediately after and before the measurement, respectively. Combining Eqs. (56) and (55), the work performed in the feedback process can be written in terms of mutual information:

$$W_{\text{fb}} \geq \Delta\mathcal{F} - \frac{1}{\beta}I(X:M). \tag{57}$$

✎

The above equation is known as the second law for feedback processes [120,126] and was experimentally verified in the motivating example discussed above [125]. It is fascinating to see how mutual information appears in the expression for feedback work, capturing the correlations established between the system and the measurement process, which lead to an increase in free energy. In particular, the examination of the thermodynamic costs inherent in the quantum acquisition of "knowledge" is a much studied field [120,121,123,126–130].

As a last observation, note that Szilard's engine can be cast in terms of the ideas discussed above. But Eq. (57) does not resolve the paradox because it constrains only the feedback stroke and therefore leaves out the energetic cost of creating the correlation $I(X:M)$ during the measurement and of resetting the demon's memory; when those contributions are added (Sec. VII B), the total work over a full cycle obeys $W_{\rm tot} \geq 0$, reinstating the second law. While this resolution is consistent with the widely adopted framework based on memory erasure, it is not the only possible perspective. Some researchers argue that erasure is not a fundamental requirement for restoring the second law. Instead, they advocate for an explicit accounting of the work and free energy involved in generating and using correlations. In this alternative approach, resetting the demon's memory to a reference state is viewed as a conventional step, not a thermodynamic necessity. This viewpoint is articulated in Ref. [131], which presents a detailed analysis of a molecular Szilard engine and argues that the apparent paradox arises from implicit assumptions rather than any deep physical necessity.

Experimental Box 3: Photonic Maxwell's demon & feedback control [132]

This experiment realizes a photonic Maxwell's demon, demonstrating that the amount of work extracted is fundamentally bounded by the information acquired through measurement, namely $|W| \propto \sqrt{I}$, with I being the single-measurement mutual information. This sublinear scaling reflects practical limitations in work extraction, while remaining consistent with the broader thermodynamic bound given by Eq. (57), which governs idealized, fluctuation-free systems.

The experiment starts with a light mode prepared in a thermal state. Measurement is performed using a high-transmittance beam splitter (BS) and avalanche photodiodes (APDs): thermal light passes through the BS, where a small fraction is reflected to a highly sensitive APD. The APD provides a binary outcome—either a click (photon detected) or no click (no photon detected). Based on the detection outcomes, the energy in each mode is inferred, and the modes are "labeled" as more or less energetic. A conditional operation (feedback) is then applied, directing these modes to two photodiodes with opposite polarities to create an energy imbalance. Both photodiodes are connected to a capacitor, which stores the extracted work by charging through this controlled energy gradient.

The two modes are then directed to photodiodes connected to a capacitor. When the feedback is applied, the capacitor charges due to the controlled imbalance. The experiment derives the following bound:

$$\frac{|\langle W\rangle|}{\sigma(W)_0} \leq \sqrt{2\langle I\rangle},$$

where $\sigma(W)_0$ quantifies fluctuations in the absence of feedback. This shows that the signal-to-noise ratio of the extracted work scales with $\sqrt{I}$.

C. Computation in the smallest nutshell

Shannon's theory provided the mathematical and conceptual foundation for understanding and optimizing how information is represented, transmitted, stored, and processed. The backbone of computing is the execution of logical operations for a given task [133]—something Shannon showed could be achieved using relays and switches. However, this architecture was slow, bulky, and prone to mechanical failure due to its moving parts. Relays consist of a coil that generates a magnetic field to open or close a mechanical contact, and this movement takes time—typically measured in milliseconds. It was soon realized that relays could be replaced by vacuum tubes, which control the flow of electrons through a vacuum using an electric field, without the need for moving mechanical components. This allowed vacuum tubes to switch states (ON and OFF) much faster than relays. In fact, they operated on the order of microseconds, as their switching time was limited only by how quickly the electric field could influence electron flow, far faster than the mechanical movement of relay parts. Beyond speed, vacuum tubes offered other advantages over relays. Relays were not suitable for signal amplification or processing analog signals. The inertia of their moving parts meant that repeatedly switching them ON and OFF (as required for fast computations) was mechanically taxing and slow. Conversely, vacuum tubes could switch rapidly between states and also amplify signals and process analog data, making them ideal for both switching circuits and more complex tasks like signal amplification and processing. However, not all challenges were resolved. The vacuum tubes generated significant heat and required large amounts of power to operate. Because they worked by heating a filament to produce electrons, this process consumed a great deal of energy and produced excessive heat.

An important element in discussing early computation is the notion of manual programming and operation. Early computers, such as those using relays or vacuum tubes, required manual programming and configuration. Their programmes were essentially hardwired into the machine until someone physically reconfigured the hardware. Switching tasks could take hours or even days, as programming involved more than just writing code—it meant altering the machine's setup itself. Once configured, the machine could automatically run tasks, but the manual setup process severely limited flexibility and made it difficult to change programmes quickly. This paradigm changed with von Neumann's 1945 report, "First draft of a report on the EDVAC" (Electronic Discrete Variable Automatic Computer) [19]. In this document, von Neumann introduced the groundbreaking concept of *stored-programme architecture*.

The central feature of von Neumann's architecture is that both the instructions (the programme) and the data are stored in the same memory. Instructions are executed sequentially, one after the other, unless the programme explicitly instructs the computer to jump to a different part of the code. This design allows computers to switch between programmes without the need for manual reconfiguration. Importantly, the memory in this system is a physical component, such as vacuum tubes, that stores both data and programme instructions, enabling the computer to dynamically access and modify the programme as needed.

Computing was in its infancy, and the main problem at the time was finding ways to make computers process data and perform operations as quickly as possible. Naively speaking, this depended on improvements in physical systems, as these are the building blocks for switching between states, from 0 to 1 and vice versa. While early engineers and scientists understood that faster computation generated more heat and that physical systems had limitations, the deeper connection between information theory and thermodynamics had not yet been fully realized.

What is the real connection between information and computation, beyond the fact that both deal with the same 0's and 1's that Shannon's information theory uses and that drive the operations inside a computer? Shannon's theory was all about measuring and transmitting information, while computation is focused on processing data using logic and instructions. As we have seen, both processing and transmitting information happens through physical systems, and they can roughly be viewed as computation tasks. Whether a computation is fast or slow depends on the physical system being used. But is computation just a mathematical abstraction, or is there a deeper, intrinsic connection between computation and information itself?

VI. THERMODYNAMICS OF COMPUTATION

What is a computer if not an engine that converts free energy into waste heat and mathematical work [25]? By the 1950s, this idea was taking shape, particularly with von Neumann's remark during a 1949 lecture [134], where he stated that a computer operating at temperature T must dissipate at least $k_B T \log 2$ of energy per elementary act of information. In this section, we discuss the fundamental thermodynamic cost of computation, specifically a lower bound of order $k_B T$ for certain data operations.

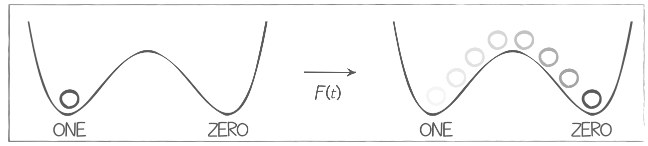


FIG. 9. *Restoring to* ZERO. A bistable potential with a particle can be recast as an information-restoring process by assuming the particle represents a bit, with its two states corresponding to the left (ONE) and right (ZERO) positions. The task consists of moving the particle from left to right using a conservative force $F(t)$ regardless of its initial position. However, such a task is not possible as this process cannot be reversed.

A. Landauer's principle

Information is physical! [22]

Despite being a short and self-explanatory statement, the above quote highlights the fact that information is stored in physical systems, such as books, hard disks, or even colorful stickers. The same systems are also used to transmit and process information. Whether through electrical or optical signals, they are inevitably bound by the laws of physics. Formally, one might ask an equivalent question: what is the interplay between performing a logical operation and its associated thermodynamic cost? This question is reinforced by the fact that information processing occurs in a physical system, and thermodynamic processes—especially beyond the thermodynamic limit (whether due to finite-time processes or finite systems)—involve energy dissipation. In his famous 1961 paper [135], Rolf Landauer relates entropy decrease and heat dissipation during logically irreversible processes, i.e., a minimum cost that must be paid to erase information.

As a warmup, one can revisit the motivating example discussed by Landauer, where the task consists of restoring a bit to a given state. We can envision a binary device as a particle in a bistable potential (see Fig. 9) and define the operation of restoring as the process that moves the particle to one of the two sites of the potential. For convenience, we assume that the left site corresponds to the state ZERO, while the right site represents the state ONE. Thus, the task is to move the particle to the state ZERO. Observe that if we know the particle's position, the task becomes simple. If the particle is already in the desired state, we do not need to do anything, and no energy is expended. However, if we know that the particle is in the ONE state, we can apply a force to push it over the barrier, and once it passes the maximum, we can apply a retarding force to slow it down. This ensures that when the particle reaches the ZERO state, it has no excess kinetic energy and no energy is expended throughout the process. Although it might be seen that restoring the particle's position to ONE can be done without expending any energy, we have actually used two different protocols depending on the particle's initial state. This leads us to ask whether a general procedure exists that can always perform the restoring action, regardless of the particle's state, without any energy cost.

In Landauer's motivating example, this question is translated into asking whether it is possible to construct a time-varying force capable of moving the particle to the ONE state, regardless of its initial position. However, since we are dealing with a conservative system, we can easily conclude that this is impossible. This becomes clear by imagining the reversed scenario. When time is reversed, the same force should work backward. This means that starting from the ONE state, the particle would retrace its steps. Since it could originally have come from either the ZERO or the ONE state, in the reversed situation, it could potentially end up in either state. This creates a paradox: starting from the ONE state, the particle would need to end up in two different places. However, according to the laws of classical mechanics, which are deterministic, a particle starting from a specific position and velocity can have only one future outcome, not two. In conclusion, the seemingly simple task of restoring a particle's position inevitably results in energy dissipation.

The motivating example can be generalized to a broader range of logical operations that have similar consequences—energy dissipation. This leads us to categorize two types of computation. A task is called *logically irreversible* if its output does not uniquely determine its inputs. In Sec. VI D, we will discuss that any logical operation that does not preserve enough information about its inputs (i.e., it is logically irreversible) implies a thermodynamic cost, such as energy dissipation. Conversely, when we can reverse the computation—such that every output can be traced back to a unique input—we call the task (or computation) reversible. The first two examples shown in Fig. 6, AND and OR, are cases where it is impossible to uniquely determine the inputs from the outputs, making them irreversible functions. In contrast, in the last example, XOR outputs 1 if exactly one of the inputs is 1. If we know one of the inputs, we can reverse the process and determine the other input from the output, making this function reversible.

To clarify the connection between logical irreversibility and entropy, Landauer revisits the example of the reset operation by imagining a collection of bits to be reset. This can be likened to a physical system where each bit is analogous to a spin in an ensemble, and the task is to align all the spins in the same direction. If the spins begin in thermal equilibrium (randomly distributed between 0 and 1) relative to a thermal reservoir at temperature T, resetting all the bits to ZERO effectively reduces the number of possible configurations the system can occupy. According to Shannon entropy [Eq. (39)], a two-state system containing a bit of information has entropy $\log 2$. After reset, the entropy is zero. Thus, this reduction corresponds to a decrease in entropy of order $k_B \log 2$ per bit. Since the total entropy of a closed system cannot decrease, this loss of entropy must be compensated elsewhere, appearing as heat dissipated into the environment. Using the Clausius inequality to express the heat flow into the reservoir, we find that the minimum heat generated during the reset process is at least

$$\beta \Delta Q_{\mathsf{E}} \geq -\Delta S_{\mathsf{S}}, \tag{58}$$

where in the present example $\Delta S_{\mathsf{S}} = \log 2$. This value represents the lower bound for energy dissipation required for the reset operation under these conditions. It is important to note that the inequality in Eq. (58) is saturated in the ideal isothermal case, which assumes an infinite time span for the process to be realized. While isothermality is an assumption needed to reach the bound, it is not what makes it unattainable. The bound becomes practically unattainable due to the requirement for an infinite duration. However, we can get remarkably close to this bound, as demonstrated in the first experimental realization in Sec. VI C. Later, in Sec. VIII D 1, we analyze the finite-time regime and explore possible corrections to this bound.

Experimental Box 4: Landauer erasure with a molecular nanomagnet [136]

This experiment realizes a Landauer erasure process in the quantum realm.

A crystal of molecular magnets (Fe_8) is used as a quantum memory, where the collective spin of each Fe_8 molecule is $S = 10$. The spin can align in one of two main orientations, $S_z = \pm 10$, corresponding to the classical bit states ZERO and ONE, making each Fe_8 molecule a qubit. The system is described by the Hamiltonian:

$$H = -DS_z^2 + E(S_x^2 - S_y^2) - g\mu_B \mathbf{S}.\mathbf{B} \tag{59}$$

where D is the anisotropy, g is the Landé g factor, and μ_B is the Bohr magneton. The first term, $DS_z^2 + E(S_x^2 + S_y^2)$, defines the energy landscape, creating a barrier that stabilizes the spin in the $S_z = \pm 10$ orientations. The third term provides external control over this landscape through the magnetic field $\mathbf{B}$.

The protocol involves three steps: (i) A magnetic field along the y axis lowers the energy barrier, allowing quantum tunneling between $S_z = +10$ and $S_z = -10$, erasing the bit by making the spin delocalized between `ZERO` and `ONE`. (ii) A magnetic field along the z axis biases the system, resetting all spins to the $S_z = +10$ (`ZERO`) state. (iii) Finally, the fields are removed, locking the spins in the $S_z = +10$ state, completing bit storage.

Therefore, logically irreversible operations, such as resetting a bit, are fundamentally tied to physical processes that result in energy dissipation. Specifically, these operations must follow thermodynamic laws, meaning that every time information is erased, a minimum amount of heat must be generated and released into the environment. This notion is now known as Landauer's principle.

B. How general is Landauer's principle?

To explain and discuss Landauer's principle, we used two specific models: the bistable potential and the spin system, each under certain assumptions that are subject to debate. A natural question arises: How general is this principle, and what are the minimal assumptions necessary to satisfy Eq. (58)? This is particularly important because there have been both theoretical [137–139] and experimental [140] reports claiming situations where Landauer's principle is apparently violated. These claims of potential violations seem to stem from the lack of a general statement formally written or rigorously proven. To address this, we revisit the work of Reeb and Wolf [141], where Landauer's principle is derived and refined using a general and minimal framework. Rather than repeating all the detailed arguments and reasoning from Ref. [141], we will present a simplified but consistent version of their proof.

We consider a minimal setup consisting of a system and a heat bath. The three main assumptions required to derive Landauer's principle are as follows. First, while the state of the system can be arbitrary, the heat bath initially starts in a thermal (Gibbs) state at inverse temperature β:

$$\gamma_{\mathsf{B}} = \frac{e^{-\beta H_{\mathsf{B}}}}{\mathrm{tr}(e^{-\beta H_{\mathsf{B}}})}, \tag{60}$$

where γ_{B} denotes the state of the heat bath, and H_{B} represents its Hamiltonian, about which we make no specific assumptions. Second, we assume that the system and the bath are initially uncorrelated

$$\rho_{\mathsf{SB}} = \rho_{\mathsf{S}} \otimes \gamma_{\mathsf{B}}. \tag{61}$$

Although the product state assumption from Eq. (61) is standard in both thermodynamics and quantum mechanics, we emphasize its crucial role in ensuring the validity of Landauer's principle. Reported violations of Landauer's bound [137,140] can be explained by recognizing that the product state assumption is not taken into account. Finally, the last assumption is that the process undergone by the system and heat bath is governed by a unitary evolution:

$$\sigma_{\mathsf{SB}} := U(\rho_{\mathsf{S}} \otimes \gamma_{\mathsf{B}})U^{\dagger}. \tag{62}$$

Importantly, the unitary assumption implies that all elements and resources are accounted for, meaning that no unspecified environment B can participate in the process or contribute to entropy.

These three assumptions are all we need to arrive at Eq. (58). Notice that while we state Landauer's principle as a consequence of the information erasure process, we have not imposed any specific constraints on our protocol (the unitary evolution) to enforce this process. This is because our goal is to formulate the problem in a way that applies to any process governed by unitary evolution, not just erasure. Erasure is simply a special case that can be related to the result we will present.

Let us begin by deriving an important relation between the initial and final entropies after the process, while introducing an information-theoretic quantity. This relation can be obtained by leveraging the properties of von Neumann entropy: additivity under the tensor product, invariance under unitary evolution, and finally, subadditivity. More precisely,

$$\begin{aligned} S(\rho_{\mathsf{S}}) + S(\gamma_{\mathsf{B}}) \overset{(18)}{=} S(\rho_{\mathsf{S}} \otimes \gamma_{\mathsf{B}}) \overset{(19)}{=} S[U(\rho_{\mathsf{S}} \otimes \gamma_{\mathsf{B}})U^{\dagger}] \\ \overset{(20)}{\leq} S(\sigma_{\mathsf{S}}) + S(\sigma_{\mathsf{B}}). \end{aligned} \tag{63}$$

Now, by manipulating Eq. (63), we can introduce and express the following quantity:

$$\begin{aligned} S(\sigma_{\mathsf{S}}) + S(\sigma_{\mathsf{B}}) - S(\rho_{\mathsf{S}} \otimes \gamma_{\mathsf{B}}) \overset{(19)}{=} S(\sigma_{\mathsf{S}}) + S(\rho_{\mathsf{S}}) - S(\sigma_{\mathsf{SB}}) \\ \overset{(21)}{=} I(\mathsf{S}:\mathsf{B})_{\sigma_{\mathsf{SB}}} \geq 0. \end{aligned} \tag{64}$$

Finally, the above equation can be further expressed as

$$-\Delta S_{\mathsf{S}} + I(\mathsf{S}:\mathsf{B})_{\sigma_{\mathsf{SB}}} \geq 0 = \Delta S_{\mathsf{B}}, \tag{65}$$

where $\Delta S_{\mathsf{X}} := S(\sigma_{\mathsf{X}}) - S(\rho_{\mathsf{X}})$ for $\mathsf{X} \in \{\mathsf{S}, \mathsf{B}\}$. Up to this point, we have merely applied the properties of von Neumann entropy to derive a relation connecting the initial and final entropies of the system and bath with the mutual information [Eq. (65)]. We will now use this relation to prove the Landauer's principle in its equality form, which reduces to the well-known Landauer bound when two negative terms are discarded.

Let us focus on the right-hand side of Eq. (65) and use the fact that one can explicitly write the term ΔS_{B} by using the assumption that the bath is initially prepared in a thermal state as given in Eq. (60). This means that one can write the entropy of the initial state as in Eq. (24): $S(\gamma_{\mathsf{B}}) = \beta \mathrm{tr}(\gamma_{\mathsf{B}} H_{\mathsf{B}}) + \mathrm{tr}(e^{-\beta H_{\mathsf{B}}})$. Note that if we add and subtract the average energy of the final state, we can write the entropy difference in terms of thermodynamic and informational quantities:

$$\begin{aligned}\Delta S_{\mathrm{B}} &= S(\sigma_{\mathrm{B}}) - \beta \mathrm{tr}(\gamma_{\mathrm{B}} H_{\mathrm{B}}) - \mathrm{tr}(e^{-\beta H_{\mathrm{B}}}) + \mathrm{tr}(\sigma_{\mathrm{B}} H_{\mathrm{B}}) \\ &\quad - \mathrm{tr}(\sigma_{\mathrm{B}} H_{\mathrm{B}}) \\ &= S(\sigma_{\mathrm{B}}) - \mathrm{tr}\left\{\sigma_{\mathrm{B}} \log\left[\frac{e^{-\beta H_{\mathrm{B}}}}{\mathrm{tr}(e^{-\beta H_{\mathrm{B}}})}\right]\right\} \\ &\quad + \beta \mathrm{tr}[H_{\mathrm{B}}(\sigma_{\mathrm{B}} - \gamma_{\mathrm{B}})] \\ &= S(\sigma_{\mathrm{B}}) - \mathrm{tr}(\sigma_{\mathrm{B}} \log \gamma_{\mathrm{B}}) + \beta \Delta Q \\ &\overset{(22)}{=} -S(\sigma_{\mathrm{B}} \| \gamma_{\mathrm{B}}) + \beta \Delta Q_{\mathrm{B}}. \end{aligned} \tag{66}$$

Finally, substituting Eq. (66) into Eq. (65) allows us to derive a general expression for the heat exchange in a unitary process, namely,

$$\beta \Delta Q_{\mathrm{B}} = -\Delta S_{\mathrm{S}} + I(\mathsf{S} : \mathsf{B})_{\sigma_{\mathrm{SB}}} + S(\sigma_{\mathrm{B}} \| \gamma_{\mathrm{B}}). \tag{67}$$

Since both mutual information and relative entropy are non-negative, discarding these two terms immediately leads to Landauer's bound.

$$\beta \Delta Q_{\mathrm{B}} \geq -\Delta S_{\mathrm{S}}. \tag{68}$$

The equality in Eq. (67) was previously derived in a different context [124], focussing on distinguishing between reversible and irreversible contributions to entropy change, but no connection to Landauer's principle was established.

As a final remark, note that Landauer's principle can naturally be written in terms of the entropy production framework since information erasure is fundamentally an irreversible process. Specifically, Eq. (68) can be cast as

$$\Sigma := \beta \Delta Q_{\mathrm{B}} + \Delta S_{\mathrm{S}} \geq 0, \tag{69}$$

where, reversible processes satisfy $\Sigma = 0$.

Finally, it is worth noting that Landauer erasure can be understood as the task of cooling a thermal state down to its ground state. According to the third law of thermodynamics, cooling a system to its ground state within a finite amount of time using finite resources is impossible. However, Landauer's bound is ideally achieved only in infinite time, thereby identifying time itself as a resource in the context of the third law [142,143]. Recently, a generalized and unified Landauer's bound was put forward in Ref. [144]. Surprisingly, finite energy and time suffice to perfectly erase information (or cool a quantum system). However, a hidden resource, termed control complexity, must diverge. This notion refers to the complexity of operations required to control the interactions between a quantum system and auxiliary systems (referred to as machines) that are specifically designed to achieve the target transformation.

✎ *Erasing with a single swap*

Consider the problem of mapping an initially maximally mixed state $\rho = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$, described by a trivial Hamiltonian $H = 0$, to a final pure state $\sigma = |0\rangle\langle 0|$. Assume that we are allowed to use a two-dimensional thermal ancilla, prepared at some inverse temperature β with energy gap E, $\gamma = (|0\rangle\langle 0| + e^{-\beta E}\,|1\rangle\langle 1|)/Z$, with $Z = 1 + e^{-\beta E}$ being its partition function. A protocol for achieving this transformation consists of applying a unitary SWAP $U_{\text{SWAP}} = |01\rangle\langle 10| + |10\rangle\langle 01| + |00\rangle\langle 00| + |11\rangle\langle 11|$, which flips both states and increases the energy gap E. However, one can easily see that the work cost for performing this protocol, i.e., $W = \Delta E_{\mathrm{B}} = [\frac{1}{2} - (e^{-\beta E}/Z)]E$, diverges as E tends to infinity. This shows that in this specific setup, perfect erasure comes at the cost of diverging energy.

✎

C. Landauer's principle in the lab

Let us roll up our sleeves and discuss the first experiment that verified Landauer's principle using a generic model of a one-bit memory [145]. In this experiment, a single colloidal particle is trapped in a double-well potential [similar to the warmup example discussed by Landauer and depicted in Fig. (9)]. Unlike the previous section, this setup is entirely described by classical mechanics and stochastic thermodynamics [26]. Therefore, assumptions such as uncorrelated states or unitary evolution, are not relevant in this context. Additionally, the notion of entropy used here is the Shannon one. Most importantly, the conclusion of this section is that, in the limit of slow erasure processes, the mean dissipated heat saturates Landauer's bound.

The setup comprises a silica bead, approximately 2 μm in diameter, trapped in an optical tweezer [146]. The double-well potential is created by rapidly alternating the focus of the laser between two distinct positions, ensuring that the particle experiences an effectively stable potential over time, despite the alternating positions. The barrier separating the two wells is high compared to the thermal energy $k_B T$, which keeps the particle trapped at one site of the well. Conversely, when the barrier is low, the particle can reach the other site. As a result, the state of the memory can be assigned a value of ZERO if the particle is in the left-hand well, or ONE if it is in the right-hand one. The memory is said to be erased when its state is reset to ZERO (or alternatively ONE), regardless of its initial state.

The experiment starts with the double well occupied with equal probability, resulting in an initial entropy of $S = k_B \log 2$. To reset the memory, the barrier height (initially larger than $8k_B T$) is lowered to $2.2k_B T$ over a period of 1 second and kept low for a time τ. Since the particle's location is uncertain, a tilting force is applied to remove the ambiguity. More precisely, the force is linearly increased up to a maximal amplitude, effectively "tilting" the landscape of the potential [see Fig. 10(a)]. The process ends by turning off the tilt and restoring the barrier to its original height, again over 1 second. The total duration of the erasure protocol is $\tau_{\text{cycle}} = \tau + 2s$.

Before turning to the quantitative results, it is useful to compare these durations with the bead's own response times. A 1-s change in the barrier is already longer than the bead needs to relax inside a single well, a fact noted by the original authors as "long compared with the relaxation time of the bead." When the barrier is low ($2.2k_\text{B}T$), the mean thermal-hopping time between wells is roughly 10 s; the erasure runs that probe Landauer's limit keep the tilt on for $\tau \simeq 25$–40 s, i.e., several hopping times, so the system remains close to equilibrium. Only the deliberately shortest protocols (with τ of a few seconds) explore a faster, more dissipative regime. It is important to note that time plays a crucial role in this process: entropy is produced, and the minimum entropy production occurs when the procedure is performed very slowly. Therefore, approaching the Landauer limit requires minimizing dissipation (and entropy production) as much as possible.

Having sketched the main ideas of the experimental protocol, let us now focus on some specific details. The tilting force is created by moving the small chamber (or "cell") that holds the single bead relative to the laser, using a piezoelectric motor. Shifting the position of the chamber changes the bead's location within the laser's trapping region, which effectively tilts the double well and guides the bead from one site to the other. The particle's trajectory, from `ZERO` to `ONE` or vice versa, is captured using a fast camera that tracks the transition during the cycle. When the state of the memory is changed, a series of double cycles is used, which moves the bead from one well to the other and back [see Fig. 10(b) for a scheme of the experimental procedure and a more detailed discussion]. In the opposite case, where the state of the memory remains unmodified, the system undergoes a reinitialization phase. This step consists of a single cycle used to ensure that the bead remains in the same well, resetting the system for the next cycle [see Fig. 10(b)]. The next question is how can we write down the explicit quantities appearing in Landauer's bound?

Since we are dealing with a microscopic system, fluctuations cannot be neglected, and as a result, thermodynamic quantities become stochastic variables. In this context, we use lowercase letters for quantities representing individual realizations along a trajectory, and uppercase letters for their averages. The dissipated heat q, along a given trajectory $x(t)$, can be derived from the first law of thermodynamics $dU = \delta w + \delta q$. For a colloidal particle trapped in a potential $U(x,t)$, the work done by the system is associated with the particle's motion, given by $\dot{x}(t)$, where $\dot{x}(t)$ is the velocity of the particle along the trajectory. The heat dissipated during this motion is related to the change in potential energy, expressed as $\delta q = -\dot{x}(t)[\partial U(x,t)/\partial x]\mathrm{d}t$. The negative sign arises because heat dissipation occurs when the system does work on the surroundings. To obtain the total dissipated

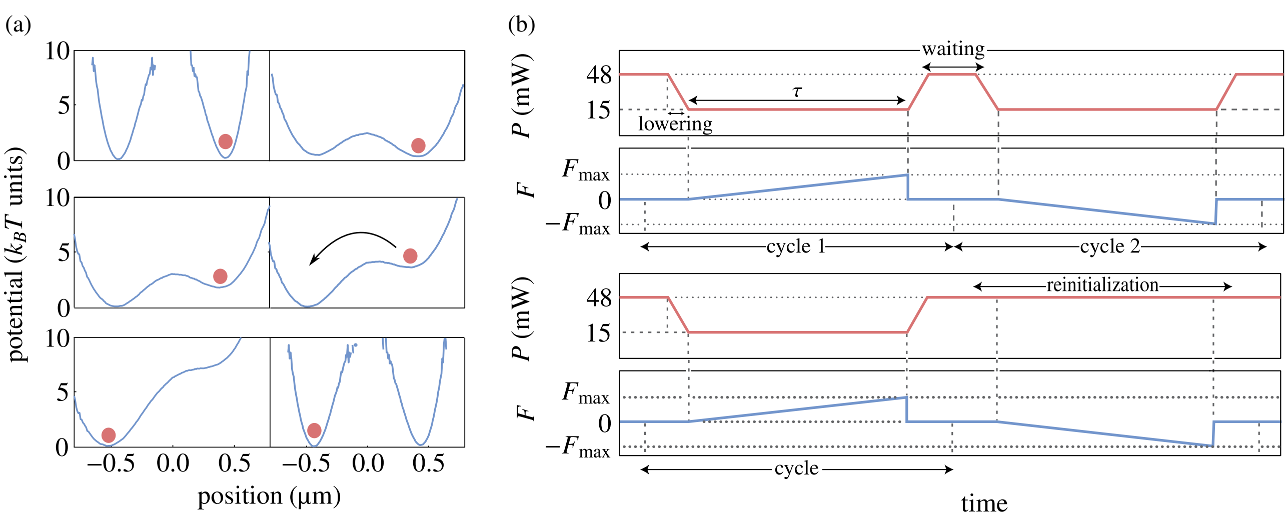


FIG. 10. *Experimental verification of Landauer's principle*. Panel (a) illustrates the erasure of one bit of information stored in a bistable potential. The process involves first lowering the central barrier, followed by applying a tilting force, which drives the particle from the left to the right. The protocol works regardless of the particle's initial position. Panel (b) (top) details the erasure protocol for when the particle transitions from 0 to 1 (or vice versa). The protocol for measuring the heat when the particle remains in the same well is depicted in the bottom of panel (b). These figures have been adapted from Ref. [145], with minor edits for consistency with this tutorial, while retaining their original meaning.

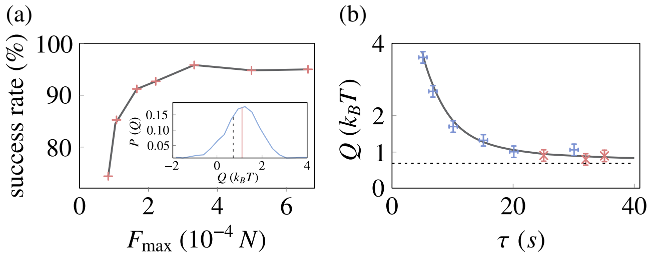


FIG. 11. *Success rate and dissipated heat*. Panel (a) shows the success rate of the erasure cycle as a function of the maximum tilt. The inset presents the heat distribution, with the solid red line indicating the average dissipated heat, and the dashed vertical line marking the Landauer bound. Panel (b) displays the average dissipated heat as a function of the protocol duration, measured for three success rates: blue for $r \geq 0.9$, red for $r \geq 0.85$, and green for $r \geq 0.75$. The horizontal dashed line represents the Landauer bound. These figures have been adapted from Ref. [145], with minor edits for consistency with this tutorial, while retaining their original meaning.

heat over a full cycle, we integrate this expression over time, resulting in the following equation for the dissipated heat:

$$q = -\int_0^{\tau_{\text{cycle}}} \mathrm{d}t \dot{x}(t) \frac{\partial U(x,t)}{\partial x}. \tag{70}$$

The average dissipated heat, obtained by averaging over all trajectories (over 600 cycles in the experiment), is always greater than the entropy difference: $Q := \langle q \rangle \geq T\Delta S = k_B T \log 2$. Reaching Landauer's bound depends primarily on two factors, the duration of the tilt τ and its maximal amplitude $F_{\max}$. If the tilt force is too weak to push the bead over the barrier, erasure will not occur [see Fig. 11(a)] for the trade-off between the success rate and the maximal amplitude). Conversely, for long durations, the mean dissipated heat does not saturate at Landauer's limit, and incomplete erasure will result in less dissipated heat. For a success rate r, the Landauer bound can be generalized as

$$Q(r) = k_B T[\log 2 + r \log r + (1-r)\log(1-r)]. \tag{71}$$

This equation shows that no heat is dissipated when $r = 1/2$, meaning the memory is left unchanged by the protocol and the transformation is quasireversible. In an ideal quasistatic erasure process ($\tau \to \infty$), the dissipated heat equals the Landauer bound. For large but finite τ, the asymptotic approach to the Landauer bound is described by $Q = Q_{\text{Landauer}} + \alpha/\tau$, where α is a positive constant [147]. For shorter times, the dissipated heat follows an exponential relaxation: $Q = Q_{\text{Landauer}} + Ae^{t/\tau_K} + \alpha/\tau$, where τ_K is the Kramers time (a characteristic time at which a Brownian particle escapes from a potential well above a potential barrier). The erasure rate and its approach to the Landauer bound are shown in Fig. 11(b).

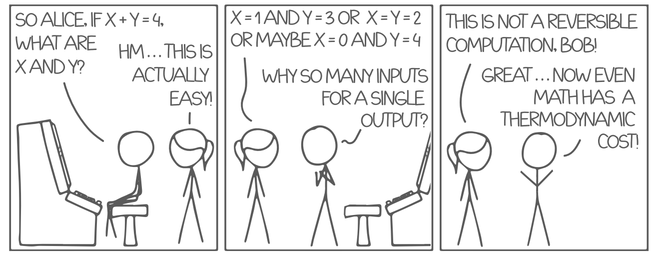


FIG. 12. *Without a history tape, no such a thing as a free lunch*. Imagine a machine that does nothing but adding numbers. If the machine outputs the number four, and Bob asks Alice to figure out the inputs that resulted in this number by running the process backward, she immediately encounters a problem. The inputs could have been one and three, or zero and four, among many other combinations. This fundamentally illustrates the concept of irreversibility: there is no single inverse solution, but rather a range of possibilities that could lead to the same outcome of four. Illustration inspired by XKCD [162].

Importantly, the attainability of the Landauer bound has been experimentally verified across various platforms. For classical systems, these include colloidal particles [145, 148–150], optomechanical systems [151] and micromechanical oscillators [152,153]. For quantum systems, the platforms range from nanomagnets [136,154,155] and superconducting devices [156] to nuclear magnetic resonance [157], ion traps [158,159] and semiconductor quantum dots [160,161].

D. Reversible computation

So far, we have hinted that typical computation is logically irreversible. But is that an unavoidable feature of computers? According to Landauer, whenever a computational task discards information about its previous state, it generates the corresponding amount of entropy. On the other hand, we might imagine that if we could somehow save all the information and steps of the computation, we could avoid this loss. For example, if we had an extra tape, initially blank, where we recorded each operation as it was performed, we could, in theory, make the computation reversible. However, as Landauer himself pointed out, this approach only postpones the problem of discarding information, especially if we plan to reuse that tape again. Consequently, a useful reversible computer would be one that, instead of storing all intermediate steps permanently, is designed to erase any unnecessary data once it is no longer needed—leaving behind only the initial input and the final output. Of course, this process still generates entropy while clearing out unnecessary intermediate information. However, the computation can still be reversed since the input and output remain intact. Surprisingly, it was shown by Bennett [24] that reversible computers that meet these requirements indeed exist. We will not reproduce Bennett's formal demonstration but instead offer a heuristic approach to support his claims. Let us begin with some basic facts: typically, computers perform operations and discard information about their history, leaving the machine in a state where it is unclear what the previous step was. When this happens, the computation is said to be logically irreversible, meaning it lacks a single-valued inverse (see Fig. 12, where Alice and Bob realize that they cannot trace back the initial steps of the computation). Notice that if we were to record each operation on a blank tape, this issue could be avoided. However, not every step needs to be saved, only those that would allow the process to be reversed, ensuring that a given output can be traced back to a unique input.

Suppose that we have a reversible computer and an initially blank tape, where every step of the computation is recorded. After the computer runs a long computation, resulting in a lengthy tape history, the question arises: can we erase the tape while preserving the final output and avoiding entropy generation?

First, by running the computation in reverse, the machine would undo each step one by one. This would eventually return the tape to its original blank state, as if the computation had never occurred. Since the forward computation was reversible, the reverse process would also be reversible. However, running the computation backward would also undo the final result, turning it back into the original input, which interferes with the purpose of performing the computation. This issue can be easily solved by making a copy of the final output on a separate tape before starting the reverse process. In this way, we can keep the output safe while reversing the rest of the computation.

During the copying process, we stop recording to the history tape to avoid generating unnecessary data. Once the output is securely copied, we can proceed with reversing the computation. While the original output will be erased during this reversal, the history tape will also be erased, restoring it to its blank state. In the end, we have the original input restored (as if nothing had been done), the final output copied preserved, and no remaining history on the tape. Even though the history tape is erased, the computation remains reversible and deterministic, as each step of the process can still be traced and reversed. This argument was formally addressed and rigorously proven by demonstrating that, given an ordinary Turing machine, a reversible three-tape Turing machine can be constructed to emulate the original on any input, while leaving behind only the input and the desired output at the end of its computation. Importantly, this argument is not limited to three-tape Turing machines but can be applied to any form of deterministic computation, whether finite or infinite, as long as it has memory to record the history.

A crucial nuance in this discussion lies in the role of the input itself. While reversible computation ensures that no entropy is generated during the processing of a given input, the preparation of that input—whether acquired from a user, a sensor, or a quantum random number generator (QRNG)—comes with its own thermodynamic costs. For instance, starting a tape with a specific input (e.g., erasing prior data to write new bits) invokes Landauer's principle, as does storing useful information for later computation. Similarly, the utility of the output matters: computations are rarely performed on random strings but on inputs meaningful to a task, and the act of selecting or generating these inputs often involves irreversible steps. Thus, while reversible algorithms avoid entropy generation during computation, the broader thermodynamic footprint of a computational task must account for the acquisition, preparation, and eventual use of inputs and outputs. This is consistent with Bennett's observation that reversible computing preserves the logical reversibility of the computation itself, but practical implementations remain subject to thermodynamic constraints at the boundaries of the system [163,164].

Bennett's technique for performing an arbitrary computation reversibly is illustrated in Table I. The process involves three tapes and three stages. The first tape, known as the *work tape*, stores the `INPUT`, intermediate steps, and `OUTPUT`. The second tape, *history tape*, logs the entire computation, recording each step taken by the machine. The third tape, *output tape*, holds the final result. The stages consist of performing the forward computation, copying the output, and reversing the computation. The detailed procedure is explained below.

Everything starts with an `INPUT` on *work tape*, which is processed through a series of computational steps and operations, eventually leading to an output. We refer to this `INPUT` processing as `WORK`. For instance, if the `INPUT` consists of numbers, the `WORK` could involve

Stage	**Tapes**		
	work tape	*history tape*	*output tape*
Forward	_INPUT	_	–
	WORK	HIST_	–
	_OUTPUT	HISTORY_	–
Copy	OUTPUT	HISTORY_	–
	_OUTPUT	HISTORY_	_OUTPUT
	_OUTPUT	HISTORY_	_OUTPUT
Reverse	_OUTPUT	HISTORY_	_OUTPUT
	WORK	HIST_	_OUTPUT
	_INPUT	–	_OUTPUT

Forward, Copy, Reverse

progression of computation

FIG. 13. *With a history tape, there is such a thing as a free lunch*. Imagine that the machine now adds and subtracts numbers. If the machine outputs the numbers four and two, Alice and Bob would immediately realize that the machine uniquely defines the inputs, making the computation reversible. Illustration inspired by XKCD [162].

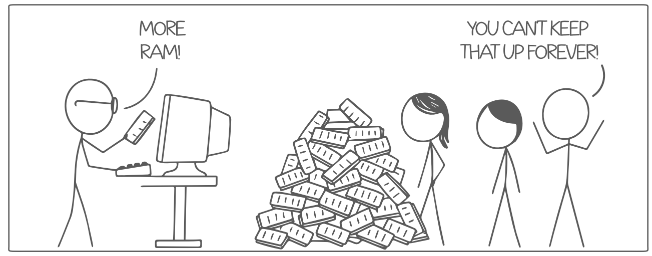


FIG. 14. *Without erasure, there is no end to lunch.*. Imagine a computer scientist determined to avoid the thermodynamic cost of erasure. Instead of deleting information, they endlessly accumulate RAM, offloading used memory into an ever-growing pile. While reversible computation avoids heat dissipation, storing every intermediate state demands increasingly more resources. Eventually, the physical burden becomes unsustainable. Illustration inspired by XKCD [162].

operations such as addition, subtraction, multiplication, or more complex logical functions. As the machine performs its computations, it records each step on the *history tape*. The machine's time steps are illustrated in Table I, where the underbar symbol indicates the position of the tape head. At the end of the forward stage, the `INPUT` has been fully processed into the `OUTPUT` on the *work tape*, while the *history tape* has captured all the steps taken during computation. In the second stage, the `OUTPUT` from the forward computation remains on the *work tape*, and the *history tape* still holds the entire computation record. The machine then copies the `OUTPUT` from the *work tape* onto the `OUTPUT` tape, ensuring the final result is stored on both tapes. Finally, in the third stage, the machine reverses the computation on the *work tape*, using the *history tape* to "uncompute" the steps. Starting at `OUTPUT`, the tape head moves backwards, undoing each step until `INPUT` is fully restored. The *history tape* is read in reverse, eventually becoming blank again. Throughout this reversal, the final result remains preserved in *output tape* (see Fig. 13).

So far, the main message of this section has been that any computation can be performed reversibly. But can we find examples of reversible or irreversible computation in nature? Quite surprisingly, the biosynthesis and breakdown of messenger RNA provide such examples. RNA synthesis, where the molecule is built step by step from DNA, is a logically reversible process—each step can, in theory, be undone without losing information (see Ref. [24] for more details). However, when cells break down RNA, they do so irreversibly, destroying the RNA without preserving any record of how it was created, resulting in the loss of information. Following Bennett's work, several studies at the intersection of computation and thermodynamics demonstrated that the second law is safe, even when considering "intelligent beings," as long as their information processing is governed by the same laws as universal Turing machines [163,165,166].

However, a critical distinction arises between reversible algorithms and physical implementations of measurement. Bennett argued that measurement itself could, in principle, be performed reversibly if the observer retains a full record of the measurement outcomes [25]. Yet, for systems like Maxwell's demon, practical constraints intervene: while the demon could theoretically measure a particle's state reversibly, it cannot indefinitely store the acquired information without eventually erasing it to reset its memory. This erasure—required to reuse the demon's finite memory—invokes Landauer's principle, which generates entropy and restores the second law. Thus, even though reversible computation avoids entropy generation by preserving intermediate steps, physical agents like the demon face unavoidable thermodynamic costs when interacting with the environment, as they cannot "unmeasure" without eventually discarding information.

Inspired by the concept of reversible computing—where computations are, in principle, reversible and all intermediate states can be recovered—one might ask whether the entropy increase from erasure could be avoided by simply accumulating memory. As illustrated in Fig. 14, this could involve endlessly adding RAM or memory modules to store every computational state. Although reversible computing may prevent heat dissipation, it demands infinite memory, which is physically unfeasible. Thus, the entropic cost, while locally or temporarily deferred, inevitably resurfaces in another form. While thermodynamic entropy in the environment might be avoided by preventing heat release, informational entropy still increases within the system—not as heat, but as mounting complexity and resource consumption embedded in the memory device's architecture.

We conclude this section by noting that similar conclusions about reversible computation were independently reached by Fredkin [167]. This became known as the billiard-ball model of computation, a prime example of a ballistic reversible computer. Conversely, the model discussed in this section is categorized as a Brownian computer, where the computation is driven by random thermal motion and energy dissipation is minimized by operating near thermal equilibrium. Ballistic computation, on the other hand, relies on the deterministic motion and collisions of particles, ideally assuming that energy is conserved through perfectly elastic collisions.

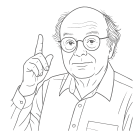
The question of how thermodynamics constrains our ability to process and manipulate information has long been a topic of exploration, approached from various perspectives [164,167–175]. Today, this remains a vibrant area of research (see Ref. [175] for a recent review). For instance, stochastic computation [176] leverages the full toolkit of stochastic thermodynamics to investigate the energetic costs of implementing computational tasks that are generally more complex than simple bit erasure [164,177,178]. Another prominent example is computation based on autonomous quantum thermal machines (see Ref. [179] for a review of quantum thermal machines). In this approach, computational tasks are encoded in the dynamics of open quantum systems, such as a few qubits interacting with multiple thermal baths [180–183].

VII. EXORCISING THE DEMON

After our grand information-computation interlude (and, of course, a historical one too), we now have all the essential tools and theoretical framework needed to resolve the paradox of Maxwell's demon. For simplicity, we will focus on Szilard's version. We begin with a less technical, heuristic (but historical) argument, then address the same problem more systematically, and conclude by presenting modern arguments for resolving this apparent paradox. The resolution is attributed to Charles Bennett [25]. However, it is important to mention that Penrose, independently and a decade before Bennett, had already pointed out that the crucial aspect in solving Maxwell's demon lay in erasing the information acquired and stored in a memory system; this would then entail an entropy cost. This is discussed in Chapter 6 of Ref. [23].

A. Heuristic approach

To start with, each cycle step is thermodynamically reversible if carried out quasistatically. This also means that the demon's measurement of the particle's position can, in principle, be performed reversibly, adding no net entropy to the universe. Although we could assume that the measurement is irreversible (which would lead to an entropy increase), our goal is to show that even if the measurement is reversible, the entropy cost will match Landauer's prediction. This reasoning echoes a crucial point stressed in Bennett's foundational review [24], where he argues that the entropy cost associated with Maxwell's demon arises not from the measurement step itself, but from the erasure or overwriting of information. Specifically, as follows:



> "This forgetting of a previous logical state, like the erasure or overwriting of a bit of intermediate data generated in the course of computation, entails a many-to-one mapping of the demon's physical state, which cannot be accomplished without a corresponding entropy increase elsewhere." [25]

This insight is directly connected to the idea of logical irreversibility discussed in Sec. VI A, where we observed that many logical operations, such as resetting a bit, correspond to many-to-one mappings, and thus, by Landauer's principle, are necessarily accompanied by entropy production. In contrast, reversible computations—where each output maps uniquely back to an input—can, in principle, be carried out without dissipation. Let us now proceed by breaking the cycle into small pieces.

Before the demon measures the particle's position, the demon's memory and the molecule's position are uncorrelated: the demon has no information about which side of the partition the molecule occupies. Once the (reversible) measurement occurs, the demon's memory and the molecule become correlated. Under perfect reversibility, this correlation does not reduce the total entropy of the demon-molecule system; rather, the measurement merely transforms their joint state from a product state to a correlated state. The crucial change is that the demon acquires knowledge of the molecule's location without any net entropy cost to (demon + molecule) at this stage.

This subtle point—that measurement can, in principle, be performed reversibly—is important. As Bennett also noted, blaming measurement for entropy production mixes up two different things: the physical act of measuring and the logical act of erasing or forgetting information:

> "...it is important...to attribute the entropy cost to logical irreversibility, rather than to measurement, because in doing the latter one is apt to jump to the erroneous conclusion that all transfers of information...have an irreducible entropy cost of order $k_B T \log 2$ per bit." [25]

In our setup, the measurement step merely establishes a correlation between the system and the memory and does so in a thermodynamically reversible way—provided it is implemented via a reversible process, such as a unitary operation on a larger system. The key is that the demon does not discard or overwrite the acquired information at this stage.

Next comes the isothermal expansion, during which the molecule is allowed to move freely in the available space. By exploiting its knowledge of the molecule's initial position, the demon can extract work during this expansion. As the molecule spreads out, its position becomes effectively randomized again, destroying the earlier correlation. In this process, the environment's entropy is reduced by one bit (since the extracted work implies less heat is dissipated), while the entropy of the demon + molecule system increases from one bit to two bits. By the end of this step, the demon's memory and the molecule appear as two independent systems, each carrying one bit of entropy: the demon retains its recorded bit, and the molecule is again equally likely to be on either side.

Finally, in order to repeat the cycle, the demon's memory must be reset to its initial blank state. According to Landauer's principle, erasing one bit of information necessarily generates at least one bit of entropy ($k_BT\log 2$) in the environment. This reset step increases the environment's entropy by one bit, canceling the reduction that occurred during the expansion. Consequently, while the demon and the molecule can be returned to their original states (with no net entropy change for demon + molecule), the environment also returns to its original entropy level. The full cycle therefore results in no overall entropy change in the demon, the molecule, or the environment, in accordance with the second law of thermodynamics.

In this way, the resolution of the paradox is tied directly to the thermodynamic cost of logically irreversible operations, as discussed in Sec. VI D. It is not the act of knowing that incurs a cost, but the act of forgetting.

B. Thermodynamics of information approach

To restore the second law in the Szilard engine, we consider the physical nature of the demon by assigning it a memory to record the measurement outcome. Operating within the classical framework (see Sec. V B), we make certain assumptions about the measurement process. First, the system and the measurement apparatus are initially uncorrelated:

$$\rho_{XY}(x,y) = \rho_X(x)\rho_Y(y). \tag{72}$$

Second, they do not interact before or after the measurement. Additionally, we assume that the measurement does not affect the system; although this is not strictly true in a quantum description, this approximation is useful in our context.

The measurement process creates a correlation between the system X and the state of the apparatus after the measurement Y'. To account for this, we update the joint state of the system and apparatus after measurement to capture these correlations. The state of the system is no longer independent because the measurement introduces new information. This leads us to the correlated state:

$$\begin{aligned}\rho_{XY'}(x,y) &= \sum_{m\in\{\mathrm{L,R}\}} \rho_X(x)\, p_{M|X}(m|x)\, p_{Y'|M}(y|m) \\ &= \rho_X(x)\left[p_{M|X}(\mathrm{L}|x)p_{Y'|M}(y|\mathrm{L}) \right.\\ &\quad \left. + p_{M|X}(\mathrm{R}|x)\, p_{Y'|M}(y|\mathrm{R})\right],\end{aligned} \tag{73}$$

where $p_{Y'|M}(y|m) = p_Y(y)/p_M(m)$ when $m = m(y)$ and is zero otherwise. Hence, for any fixed apparatus readout y only the term with $m = m(y)$ is nonzero. We label the two possible measurement outcomes by L (particle in the left half of the box) and R (particle in the right half). Note that the marginal density $\rho_X(x)$ is unchanged by the measurement, while the apparatus state changes from Y to Y'.

Since the energy of the system does not change during the measurement, the nonequilibrium free energy of the global system after the measurement can be written as

$$\mathcal{F}(X,Y') = \mathcal{F}(X) + \mathcal{F}(Y') + \frac{1}{\beta}I(X:M). \tag{74}$$

Consequently, the work needed to perform the measurement satisfies

$$W_{\mathrm{meas}} \geq \Delta\mathcal{F}_{\mathrm{tot}} = \Delta\mathcal{F}_Y + \frac{1}{\beta}I(X:M), \tag{75}$$

where $\Delta\mathcal{F}_Y = \mathcal{F}(Y') - \mathcal{F}(Y)$. Since $I(X:M)\geq 0$, creating correlations between the two subsystems increases the free energy. If this increase is not balanced by a decrease in $\Delta\mathcal{F}_Y$, work must be supplied, and heat dissipation occurs.

The demon can extract work $W_{\mathrm{ext}} \geq (1/\beta)I(X:M)$ using the information acquired during measurement in a cyclic process where the system is returned to its initial state X. However, to complete the cycle, the apparatus must also be returned to its initial state Y. Therefore, the demon must perform work to reset the apparatus, given by $W_{\mathrm{reset}} \geq \mathcal{F}(Y) - \mathcal{F}(Y') = -\Delta\mathcal{F}_Y$. As a result, the total work involved in the process is

$$W_{\mathrm{tot}} = W_{\mathrm{meas}} + W_{\mathrm{fb}} + W_{\mathrm{reset}}, \tag{76}$$

where $W_{\mathrm{fb}} = -W_{\mathrm{ext}}$ is the work extracted during the feedback process. Therefore, the validity of the second law for feedback processes is restored when the entropy costs of measurement and resetting the demon's memory are taken into account. This generalizes Bennett's analysis of the Szilard engine. Bennett discussed the case where $W_{\mathrm{meas}} = 0$, and the demon must overwrite the outcome of the measurement, performing work $k_BT\ln 2$. This situation corresponds to $\Delta\mathcal{F}_Y = -TI(X:M)$.

From the previous discussion, we learn that the Szilard engine can be framed in terms of a simple exchange between work and the free energy stored in the correlations between the system and the demon. For example, if $\Delta\mathcal{F}_Y = 0$, the engine operates by creating correlations during measurement, an operation that requires work $TI(X : M)$ and increases the free energy by the same amount, and then destroying these correlations during feedback, where the same amount of work is extracted.

C. General approach

We now turn to a more abstract and modern approach. We aim to capture a scenario in which the system begins in a given (unknown) state, and the demon measures the system, using the acquired information to extract work, all while ensuring that the system is returned to its original state.

The setup consists of a system S, a thermal environment E, a memory system M, and a work reservoir W. While we do not make any assumptions about the initial state of the system or the memory, we assume that the environment acts as an ideal reservoir (in the classical sense). In addition, the work reservoir can be associated with a specific system in a particular state. This follows from the physical nature of work: when work is performed, it results in changes to the state of a physical system. Work must be stored in some form, such as the potential energy of a weight, the charge state of a battery, or other systems capable of storing energy. Hence, the work reservoir represents a physical system whose state encodes the amount of work performed, and this state is described by a density operator. This formalism also allows us to distinguish between two contributions: the work that stores energy without increasing entropy (zero entropic cost) and the part associated with maximum entropic cost (the thermal reservoir).

Each subsystem is described by a density operator ρ_{X} with $\mathsf{X} \in \{\mathsf{S}, \mathsf{E}, \mathsf{M}, \mathsf{W}\}$. We assume that the composite system is closed and evolves via an energy-preserving unitary U as

$$\sigma_{\mathsf{SEMW}} := U(\rho_{\mathsf{S}} \otimes \rho_{\mathsf{E}} \otimes \rho_{\mathsf{M}} \otimes \rho_{\mathsf{W}})U^{\dagger}. \tag{77}$$

Using the subadditivity of entropy, one can write the following inequality:

$$S(\sigma_{\mathsf{S}}) + S(\sigma_{\mathsf{E}}) + S(\sigma_{\mathsf{W}}) + S(\sigma_{\mathsf{M}}) \geq S(\sigma_{\mathsf{SEMW}}). \tag{78}$$

Since the von Neumann entropy is invariant under unitary transformations, we can replace $S(\sigma_{\mathsf{SEMW}})$ by $S(\rho_{\mathsf{SEMW}})$. Then, using the fact that the initial state is uncorrelated, we have $S(\rho_{\mathsf{SEMW}}) = S(\rho_{\mathsf{S}}) + S(\rho_{\mathsf{E}}) + S(\rho_{\mathsf{W}}) + S(\rho_{\mathsf{M}})$. This leads to the following inequality:

$$\Delta S_{\mathsf{S}} + \Delta S_{\mathsf{E}} + \Delta S_{\mathsf{W}} + \Delta S_{\mathsf{M}} \geq 0. \tag{79}$$

Because the protocol brings the system back to its initial state, its entropy change vanishes, $\Delta S_{\mathsf{S}} = 0$. The work reservoir is modeled as an ideal deterministic battery: at every moment, it occupies a definite energy eigenstate—e.g., a weight at a well-defined height—so its state is given by $\rho_{\mathsf{W}}(t) = |E(t)\rangle\langle E(t)|$, and its entropy satisfies $S[\rho_{\mathsf{W}}(t)] = 0$ for all t. This implies $\Delta S_{\mathsf{W}} = 0$. Given that we are dealing with an ideal thermal reservoir, the change in its entropy is related to the exchanged heat as $\Delta S_{\mathsf{E}} = \beta Q_{\mathsf{E}}$. Combining this with the entropy balance in Eq. (79), and using $\Delta S_{\mathsf{S}} = \Delta S_{\mathsf{W}} = 0$ as argued above, we obtain

$$\Delta S_{\mathsf{M}} \geq -\beta Q_{\mathsf{E}}. \tag{80}$$

This inequality shows that the amount of heat extracted from the bath and converted into work must be compensated by a corresponding increase in the entropy of the memory system.

VIII. WHAT'S NEXT?

Maxwell's demon was resolved back in the 1980s, but its implications continue to inspire novel theoretical frameworks that incorporate information into thermodynamics [59,184]. The formalism for studying thermodynamics in both classical and quantum systems—especially in situations beyond equilibrium, where fluctuations are significant—is known as (quantum) stochastic thermodynamics [26,185] and quantum thermodynamics [27–33]. In what follows, we bring a list (though not exhaustive) timeline of active areas influenced by Maxwell's demon. At the beginning of each section, we also mention the essential theoretical tools for readers who want to explore these topics further.

A. Physical models and feedback control

Could an "authentic" Maxwell's demon be more than just an agent sitting in a gas chamber, selectively sorting particles? Although Maxwell's demon was originally conceived within a paradigm where it selectively sorts gas molecules based on their velocity, we discussed earlier how it can also be viewed as an information-gathering and processing device that seemingly extracts work at no cost. Consequently, various physical implementations of Maxwell's demon, both classical and quantum, have emerged in recent years [185–209]. They used different implementations to provide a complete thermodynamic description of a Maxwell demon model and the system on which it acts. However, as extensively discussed here, it is not possible to violate the second law of thermodynamics using a Maxwell's demon. Nonetheless, they can still be interpreted as a feedback system. That is, the demon observes the system, gathers information about its microscopic states, and uses this information to perform a thermodynamic process [120,122,126,185,210–217].

As an example of both the physical implementation of a Maxwell demon and the thermodynamic behaviour of feedback-controlled systems, we briefly mention the results from Ref. [185]. In this work, a demon reduces the apparent entropy of a subsystem by processing information and influencing the electron flow. However, this reduction in entropy comes at a cost—the demon must dissipate energy and produce entropy to exert its influence, ensuring that the total entropy production of the entire system remains non-negative. This model demonstrates how feedback control and information flow can be incorporated into nonequilibrium thermodynamics, providing insights into the role of information in modifying thermodynamic processes.

The model consists of two single-level quantum dots, one of them being the main system S and the other acting as the demon D. Both are described by fermionic annihilation and creation operators $c_{\mathsf{X}}/c_{\mathsf{X}}^{\dagger}$ and respective energy ϵ_{X}, where $\mathsf{X} \in \{\mathsf{S}, \mathsf{D}\}$. Both quantum dots interact via Coulomb repulsion U. The Hamiltonian describing this interaction is given by

$$H = \epsilon_{\mathsf{S}} c_{\mathsf{S}}^{\dagger} c_{\mathsf{S}} + \epsilon_D c_{\mathsf{D}}^{\dagger} c_{\mathsf{D}} + U c_{\mathsf{D}}^{\dagger} c_{\mathsf{D}} c_{\mathsf{S}}^{\dagger} c_{\mathsf{S}}. \tag{81}$$

The system dot S is connected to two heat baths L and R, which have the same temperature but different chemical potentials. This creates a flow of electrons between the reservoirs through the dot, generating a particle current. This setup constitutes a single-electron transistor (SET). The demon dot D is coupled to a separate reservoir and is also capacitively coupled to the system dot. This means that dot D can sense and react to whether dot S is filled or empty—much like the original idea of Maxwell's demon being able to "see" and act on individual particles.

The dynamics of the system are described using a Markovian master equation, which calculates the probabilities of electrons moving between different states in the dots over time. The figure of merit here is entropy production, which remains non-negative throughout the process. However, an interesting effect emerges: the demon alters the amount of entropy produced by manipulating the flow of electrons through the dot S. More precisely, the demon's ability to monitor and respond to the state of the system dot effectively lowers the apparent entropy of dot S. This entropy reduction comes at a cost—dissipation in the demon dot and its reservoir compensates for the reduced entropy in dot S, ensuring the total entropy production remains non-negative.

Several proposals for nanoelectronic circuits have been put forward since Ref. [185]—see Review [218] for recent experiments on quantum heat transport, fluctuation relations, and implementations of Maxwell's demon—and have been implemented experimentally in Refs. [87,219]. Beyond the single-electron transistor, there are many other ways to implement a Maxwell demon and explore feedback mechanisms.

Importantly, the model discussed above is closely related to an earlier model by Sánchez and Büttiker [220], where a similar mechanism was studied in the context of power generation. While the fundamental equations governing the system are the same, the explicit interpretation in terms of information exchange and feedback was developed in later works [221–224]. Experimental realizations of similar setups have also been reported [225,226], with connections to the Feynman ratchet problem [227]. However, a crucial limitation of these models is that, while they exhibit apparent second-law violations in the conductor, they also violate the first law, making them not "strict" Maxwell demons. A more refined version was later proposed [228], where all demonic actions—including opening and closing gates—are explicitly incorporated. The criteria for identifying true Maxwellian demons have been further developed in Ref [229], including considerations of internal current reversal and cross-correlations, recently explored in coupled qutrit systems [230].

Experimental Box 5: Rectifying entropy production on Maxwell's demon [231]

This experiment realizes a Maxwell's demon in the form of a feedback control mechanism.

The system, a spin-$\frac{1}{2}$ represented by a carbon nucleus (^{13}C), is initially described by the Hamiltonian $H_0 = \frac{1}{2}\hbar\omega\sigma_z$ and is prepared in a thermal state. The memory, modeled by a hydrogen nucleus (^{1}H), starts in its ground state. The protocol involves three main steps. (i) The Hamiltonian of the carbon nucleus changes rapidly from H_0 to $H_\tau = (\hbar\omega/2)\sigma_x$ by applying an external magnetic field. This quick change drives the system out of equilibrium. (ii) To establish a correlation between the system and the memory, a CNOT operation is applied, which effectively "links" the states of the carbon and hydrogen nuclei. Next, a projective measurement is performed on the carbon nucleus, which captures information about its state. (iii) Based on the measurement result, a controlled operation is applied to the system. This operation is chosen based on the outcome of the measurement and allows the protocol to control the production of entropy by directing the system evolution in a targeted way.

Continuous feedback

Recently, variants of the original Maxwell demon that operate continuously in time have emerged, based on the concept of continuous quantum feedback [232,233]. Unlike conventional feedback control, these demons continuously monitor the system and use the acquired information to perform a thermodynamic process.

A beautiful illustration of this idea is the concept of gambling demons introduced in [234]. Roughly speaking, the demon invests work into a system to perform a thermodynamic process, guided by a gambling strategy. At various points during the process, it must decide whether to continue investing more work or stop and "cash out" the remaining work. Since the system's evolution is stochastic, its future behavior remains inherently unpredictable. Remarkably, in specific gambling schemes, the demon can, on average, extract more free energy than the work invested over many iterations—a scenario forbidden by the standard second-law inequality [Eq. (49)]. However, by accounting for the information acquired during the process and the nature of the protocol, a generalized second-law-like inequality is derived.

The setup consists of a thermodynamic system, which may be in equilibrium or out of equilibrium, interacting with a inverse temperature heat bath β. The system is characterized by a Hamiltonian $H[\lambda(t)]$, which depends on an external parameter $\lambda(t)$, and its state is represented by a probability density $\rho(x,t)$, with x denoting a microstate and $x_{[0,t]} = \{x(t)\}_{t=0}^{\tau}$ a given trajectory. The protocol involves varying deterministically the external parameter $\lambda(t)$ with a total duration of τ. The system's evolution is described within the framework of stochastic thermodynamics, where thermodynamic quantities are expressed as functionals of the stochastic trajectory $x_{[0,\tau]}$. The gambling strategy is defined via a generic stopping condition that depends on the information collected about the system up to the current time. In each run, the demon gambles by applying this prescribed stopping condition. Since the demon must decide to stop before or at the end of the nonequilibrium drive, stopping times satisfy $\mathcal{T}(x_{[0,\tau]}) \leq \tau$ for any trajectory $x_{[0,\tau]}$. For these systems, the inequality Eq. (10)] is generalized to

$$\langle W\rangle_{\mathcal{T}} - \langle \Delta F\rangle_{\mathcal{T}} \geq -k_B T\langle \delta\rangle_{\mathcal{T}}, \tag{82}$$

where the average is over taken over many trajectories $x_{[0,\mathcal{T}]}$, each stopped at a stochastic time $\mathcal{T}$. The term on the right-hand side of Eq. (82) is the stochastic distinguishability, defined as

$$\delta(\mathcal{T}) := \ln\left[\frac{\varrho(x(\mathcal{T}),\mathcal{T})}{\tilde{\varrho}(x(\mathcal{T}),\tau-\mathcal{T})}\right], \tag{83}$$

where $\varrho(x(\mathcal{T}),\mathcal{T})$ is the probability density of the system's state at the stopping time $\mathcal{T}$ in the forward process, and $\tilde{\varrho}(x(\mathcal{T}),\tau-\mathcal{T})$ corresponds to the probability density at the same stopping time in the reference time-reversed process. This term captures how distinguishable the forward and reverse trajectories are at stopping time. This leads to the following generalized fluctuation theorem, $\langle e^{-\beta(W-\Delta F)-\delta}\rangle_{\mathcal{T}} = 1$ [compare it with Eq. (15)]. Importantly, the idea of a gambling demon can be extended to the quantum realm by considering quantum jump trajectories. The key difference in this case is that the fluctuation theorem derived earlier now includes an additional entropic term associated with the quantum measurement process.

Experimental Box 6: Gambling single-electron box [234]

This experiment realizes a gambling demon and verifies the modified second-law inequality with a 99.5% of fidelity.

The main system consists of two metallic islands connected by a tunnel junction, forming a single-electron box (SEB). At low temperatures, the SEB can be approximated as a two-level system with charge number states $n = 0$ and $n = 1$. The system is driven by an external gate voltage V_g, which controls the offset charge n_g. The tunneling of an electron between the islands corresponds to transitions between the states $n = 0$ and $n = 1$, determined by n_g and associated with an energy cost. The system is continuously monitored by a SET, which is capable of distinguishing small changes in the electrostatic potential of the box caused by the presence or absence of an extra electron. This real-time monitoring allows for precise tracking of the system's state.

The protocol begins with the system in thermal equilibrium, where the initial energies are uniformly distributed. The energy split between the states is then tuned by controlling n_g. By continuously measuring the system's charge state, the SET provides the information needed to reconstruct the stochastic trajectory of the system. The protocol is repeated multiple times to gather sufficient statistics. The gambling strategy involves stopping the dynamics at stochastic times when the work exceeds a predefined threshold. This strategy leverages the continuous monitoring and feedback provided by the SET to apply a stopping rule dynamically.

Maxwell's demon with continuous quantum feedback control has also been explored in the context of generating many-body entanglement in a main system. For instance, the demon operates by randomly selecting two qubits from the system using a roulette mechanism. It then implements quantum feedback control on the selected qubits, simultaneously reducing entropy and enhancing correlations between them. By continuously repeating this process of selection and feedback control, entanglement of many bodies can be generated [235]. Beyond the two examples discussed here, the incorporation of continuous feedback has become a very active area of research [200,221,236].

B. Maxwell's demon and quantum properties

Quantum mechanics introduces a range of features that are absent in classical systems. We have already explored some of these, such as the role of particle statistics in the Szilárd engine, and briefly mentioned that Maxwell's demon and continuous feedback control can generate entanglement. But what thermodynamically happens when Maxwell's demon interacts with a system that exhibits uniquely quantum properties, such as entanglement and coherence?

In the early 2000s, results exploring the interplay between work extraction and quantum correlations within a Maxwell's demon scenario began to emerge [237–239]. First, thermodynamic inequalities were derived to distinguish entangled states from classically correlated ones based on the amount of extractable work, leading to a work extraction protocol that acts as a separability criterion [239]. Second, the concept of efficiency between quantum and classical demons was introduced, with the difference in the extraction capabilities of the work quantified by quantum discord, a measure of "quantumness" in correlations [238]. This efficiency difference demonstrated that quantum demons, by exploiting quantum correlations, could extract more work than classical demons, thus providing a thermodynamic advantage and complementing previous findings on the unique role of entanglement in work extraction. These later led to a number of works underlying the thermodynamics of correlations [240–251]

Extending Maxwell's demon to the quantum domain, or linking thermodynamic quantities and quantum features, often relies on assuming specific models for both the system and the demon's memory. However, a general and minimal setup that addresses both of these issues was proposed in Ref. [252]. The authors considered a minimal configuration comprising a quantum system, a quantum memory, and a thermal environment. The main system is unknown and arbitrary, and while the initial state of the memory is also arbitrary, it is required that it begins and ends in the same state. This ensures that the memory affects the system's dynamics without changing its energetics, so heat exchange occurs only between the system and its environment. As for the environment, no restriction on its dimension is imposed, other than that it was initially prepared in a thermal Gibbs state at temperature T. The composite system is assumed to be closed and evolves under an energy-preserving unitary [253]. Without additional constraints, no assumptions are made about the strength of the interaction (whether weak or strong), its complexity (whether local or collective), or its duration (short or long) relative to natural time scales. As there is no additional source of energy or battery system, energy exchange between the main system and the environment is accounted for as heat.

Within this minimal setup, fundamental bounds on the exchange between the quantum system and the thermal environment led to a direct correspondence between thermodynamic and quantum features. Surprisingly, these results also revealed that quantum properties can be detected by monitoring heat exchange in a quantum process.

C. Maxwell's demon and quantum heat engines

Quantum thermodynamics, an emerging field attempting to translate thermodynamics laws and understand how thermodynamic process are carried out in the quantum realm, embraced Maxwell's demon to itself. One of the first papers in the field suggested that a three-level maser could be regarded a heat engine [254]. It did not take long after the study of quantum heat engines was formalized [255,256] for Maxwell's demon to stir the pot.

1. Maxwell's demon as an engine

The analogy between Maxwell's demon and a quantum heat engine was first drawn in Ref. [257], where it was emphasized that a quantum demon is nothing more than an interaction between two quantum systems that enables the controlled transfer of information from one to the other—essentially making it an information-processing quantum heat engine.

In fact, a quantum heat engine can effectively be seen as the demon itself, with their operations emulating a Maxwell's demon. This was discussed by Kieu [258,259], using a two-level system undergoing quantum adiabatic processes and energy exchanges with heat baths. By using quantum measurement and control processes to selectively transfer energy, the machine's functionality resembles the behavior of Maxwell's demon, sorting molecules based on their temperature [260]. Different models of heat engines implementing Maxwell's demon or the Szilárd engine also emerged [261–263].

2. Maxwell's demon assisting an engine

What if, instead of assuming that the heat engine is a demon per se, we allow a demon to assist a quantum heat engine? This problem was introduced by Quan *et al.* [264], who proposed a new quantum heat engine model with a built-in quantum Maxwell demon that performs both the quantum measurement on the working substance and a feedback control for the system according to the measurement. This work led to two important concepts that continue to be actively explored today: quantum heat engines assisted by Maxwell's demons [265–269] and the use of the demon as feedback control.

D. Landauer's principle

The past decade has provided significant insight into Landauer's principle, especially in far-from-equilibrium situations [121,157,172,188,270–298]. In what follows, we briefly comment on some of the progress relating to

Landauer's principle (see Ref. [299] for a recent review). The theoretical tools typically used are varied, ranging from the resource-theoretic approach to quantum thermodynamics [60,61,300,301], to notions of entropy production [43], and the framework of thermodynamic geometry [302,303].

1. Finite-time Landauer erasure

The Landauer bound is optimally achieved in the quasistatic limit, where the erasure process is carried out very slowly, minimizing dissipation. This naturally raises the question of what happens when the erasure process is performed within a finite time. For classical systems governed by overdamped Langevin dynamics, optimal erasure protocols have been derived for slow but finite-speed processes [276,278,279] and were later generalized to allow for arbitrary driving speeds [289,291]. For quantum systems described by Markovian open quantum dynamics, the optimal erasure cost, valid for arbitrary operational times, was found [292,294]. Importantly, in Markovian open quantum systems, quantum coherence has been shown to incur additional heat costs during erasure. However, this conclusion is context dependent: recent work demonstrates that coherence can play a constructive role in optimizing thermodynamic costs for certain protocols, particularly in non-Markovian settings [304]. The interplay between coherence and erasure efficiency remains nuanced, as its utility or detrimentality depends on the driving regime and system-environment coupling. For slowly driven unitary quantum systems, it remains an open question whether coherence is necessary or unnecessary for minimising dissipation. The take-home message is that the minimum dissipation for erasing a classical bit has a lower bound set by the Landauer cost, plus an additional term that scales inversely with the operational time:

$$\Delta Q_{\mathrm{B}} \geq k_B T \log 2 + \frac{\alpha}{\tau}, \tag{84}$$

where τ represents the total time of the process, and $\alpha \in \mathbb{R}$ is a positive constant that depends on the specific process and the nature of the system (whether classical or quantum).

Experimental Box 7: Finite-time Landauer erasure in a quantum dot [161]

This experiment realizes Landauer erasure with nonlinear protocols that minimize dissipation beyond linear approaches.

The setup consists of three quantum dots (D_1, D_2, D_3) on an InAs nanowire. The main dot, D_1, holds the bit of information in its occupancy state, which can be 0 (unoccupied) or 1 (occupied). This state represents the bit to be erased. The energy of D_1 is controlled by a gate voltage, which allows for precise manipulation during the erasure process. The dot D_2 is prepared in a state that effectively acts as a heat bath for D_1. Finally, D_3 functions as a sensor, detecting changes in occupancy in D_1 and enabling real-time monitoring.

The erasure protocol begins by allowing D_1 to reach thermal equilibrium at an initial energy level, E_0, where it has a 50% chance of being occupied. The energy is then gradually increased to E_1, making D_1 almost certainly unoccupied and thus erasing the bit. After this, the energy is quickly reset to E_0. Heat dissipation during this process is measured by tracking electron transitions. Unlike traditional linear approaches, nonlinear protocols that minimize dissipation are employed.

Using tools from thermodynamic length, this work provides the first experimental demonstration of minimising dissipation through nonlinear driving in a quantum dot.

If in the protocol, the stored information is fully erased, then $\alpha = \bar{\gamma}_\tau^{-1}$, where $\bar{\gamma}_\tau := \tau^{-1}\int_0^\tau \sum_k \mathrm{tr}\{L_k(t)\rho_t L_k(t)\}\mathrm{d}t$ is the time average characterizing the thermal relaxation timescale, $L_k(t)$ denotes the time-dependent jump operators and ρ_t the state of the system to be erased. Importantly, the aforementioned studies focused on Markovian systems that weakly interact with a heat bath. Under strong coupling conditions, however, one may question how the constant α in Eq. (84) looks like, as new effects from faster relaxation rates and non-Markovian dynamics come into play. This question was addressed by modeling a bit (a two-level system) encoded in the occupation of a single fermionic mode strongly interacting with a heat bath [305]. Specifically, $\alpha = k_B T a \tau_{\mathrm{Pl}}$, where $a \approx 2.57946$ and $\tau_{\mathrm{Pl}} = \hbar/k_B T$ represent the Planckian time [306]. In particular, this finite-time correction incorporates the product of two fundamental constants of nature—Boltzmann's constant k_B and Planck's constant $\hbar$.

2. Landauer's principle at zero temperature

Upon a closer look at the Landauer bound given by Eq. (58), we note that it becomes trivial in the zero-temperature limit. This result can be physically understood by noting that as $T \to 0$, the heat bath approaches its ground state, meaning that any physical process must satisfy $\Delta Q_{\mathrm{B}} \geq 0$. The bound is useful in showing that certain processes can occur with zero heat cost, but beyond this, it provides no further information. This motivates the question of whether it is possible to derive a modified bound that captures nontrivial information even at zero temperature. Very surprisingly, using minimal assumptions—specifically, that the environment is in a thermal state—a tighter bound was derived in Ref. [286]. More specifically,

$$\Delta Q_{\mathsf{B}} \geq \mathcal{Q}[\mathcal{S}^{-1}(\Delta S_{\mathsf{S}})], \tag{85}$$

where $\mathcal{Q}(\beta') := \mathrm{tr}\{H_{\mathsf{E}}[\gamma_{\mathsf{E}}(\beta') - \gamma_{\mathsf{E}}(\beta)]\}$ and $\mathcal{S}(\beta') := S[\gamma_{\mathsf{E}}(\beta')] - S[\gamma_{\mathsf{E}}(\beta)]$, with $\gamma_{\mathsf{E}}(x)$ representing the thermal state of the heat bath at a given inverse temperature x. Observe that the tricky aspect lies in the fact that $\mathcal{Q}(\beta')$ is monotonically decreasing in β' and thus has a unique inverse $\beta' = \mathcal{Q}^{-1}(\Delta Q_{\mathsf{E}})$. The difference now is that the bound relates less trivial functions; however, it involves only thermal equilibrium quantities, despite the process being arbitrarily far from equilibrium.

To make the bound more concrete and suitable for calculations, it is helpful to express $\mathcal{Q}$ and $\mathcal{S}$ in terms of the environment's heat capacity. For a bath at temperature T with heat capacity $C_{\mathsf{E}}(T)$, one finds [286]:

$$\mathcal{Q}(T') = \int_T^{T'} C_{\mathsf{E}} \mathrm{d}\tau, \quad \mathcal{S}(T') = \int_T^{T'} \frac{C_{\mathsf{E}}(\tau)}{\tau} \mathrm{d}\tau, \tag{86}$$

These expressions allow the modified bound in Eq. (85) to be evaluated directly, provided a model for the bath is specified.

We now illustrate how Eq. (85) can be applied by considering a simple yet instructive example: a qubit system initially in a maximally mixed state and undergoing an erasure process, while coupled to a low-temperature phononlike environment with heat capacity $C_{\mathsf{E}}(T) = aT^3$, where a is a positive constant [307]. Using Eq. (86), we obtain

$$\mathcal{S}(\beta') = \frac{a}{3}(T'^3 - T^3), \quad \mathcal{Q}(\beta') = \frac{a}{4}(T'^4 - T^4). \tag{87}$$

In the zero-temperature limit $T \to 0$, the entropy change simplifies to $\mathcal{S}(\beta') = (a/3)T'^3 = \Delta S_{\mathsf{S}}$. Solving for T' gives

$$T' = \left(\frac{3\Delta S_{\mathsf{S}}}{a}\right)^{1/3}. \tag{88}$$

Substituting this into $\mathcal{Q}$ yields the lower bound on the heat:

$$\Delta Q_{\mathsf{B}} \geq \mathcal{Q}[\mathcal{S}^{-1}(\Delta S_{\mathsf{S}})] = \frac{a}{4}\left(\frac{3\Delta S_{\mathsf{S}}}{a}\right)^{4/3} = \frac{3^{4/3}}{4}\frac{(\Delta S_{\mathsf{S}})^{4/3}}{a^{1/3}}. \tag{89}$$

This result shows that, even at zero temperature, erasing one bit of information ($\Delta S_{\mathsf{S}} = \log 2$) requires a strictly positive amount of heat to be dissipated, provided the environment has a nontrivial heat capacity. The bound remains meaningful in regimes where the standard Landauer principle becomes uninformative.

This new bound can be used to study a variety of systems, such as the paradigmatic example of the spontaneous emission of a two-level atom into a single-mode cavity, the interaction of a system with a one-dimensional waveguide, and many other examples (see Ref. [286]).

3. Finite-size corrections to the Landauer erasure

One may also ask: what are the finite-size corrections to Landauer's principle when the heat bath has a finite number of particles? More precisely, consider a reservoir consisting of n particles, prepared in a Gibbs state at an inverse temperature β, and coupled to a quantum system. Can we derive bounds on the thermodynamic energetic cost in this scenario? This question naturally leads to an additional consideration of whether the bath consists of noninteracting particles or includes interactions among its components. For any thermodynamic process, with the help of an environment with a noninteracting Hamiltonian, the entropy production satisfies [308]

$$\Sigma \geq \left(\frac{\Delta S_{\mathsf{S}}}{\log d}\right)^2 \frac{1}{n}. \tag{90}$$

That is, it decays at most as $\propto 1/n$ in the presence of a noninteracting environment. For a single qubit, the above result becomes $\Sigma \geq 1/3n$. Several works have analyzed the decay of Σ with n [113,309–311], finding a convergence of the form $\Sigma = A/n$ for $n \gg 1$, where A depends on the specific protocol. A recent work derived the "best" (known) protocol for a class of collisional models, finding $A \approx \pi^2/8$ [311].

Observe that Eq. (90) holds for thermal environments with noninteracting Hamiltonian. We can ask whether it can be violated in the case of interacting Hamiltonians. If so, this would demonstrate an advantage over noninteracting ones. To answer this question, we start by noting that for sufficiently large n, a finite-size correction that is valid in *any* environment for thermodynamic processes that reduce the system's entropy $\Delta S_{\mathsf{S}} < 0$ is given by [309]

$$\Sigma \geq \frac{2(\Delta S_{\mathsf{S}})^2}{\log^2(d-1) + 4} = \mathcal{O}(n^2). \tag{91}$$

Equation (91) holds universally, regardless of the specific system-environment interaction and the nature of the process being implemented. The achievability of Eq. (91) was recently proved with an explicit process that realizes Landauer erasure considering heat baths with interacting Hamiltonian [308]. Specifically, the authors considered an initial state $\rho_{\mathsf{S}} = \mathbb{1}_{\mathsf{S}}/2$ and constructed a unitary operator and a Hamiltonian H_{B}, which mapped ρ_{S} to a state σ_{S}, which is ϵ close to the ground state. In doing so, they found that the entropy production in this protocol is bounded by

$$\Sigma \geq 2\left(\frac{\pi}{n}\right)^2 + \mathcal{O}(1/n^2). \tag{92}$$

Therefore, we see that entropy production decays quadratically with the size of the environment. This is in stark contrast to the case of noninteracting environments, where entropy production can decrease at most linearly with the size of the environment.

4. Optimal cost of erasure in the single-shot scenario

Another extension of Landauer's erasure principle considers the protocol within the so-called single-shot scenario [60,312–315]. This setting removes the assumption of having either infinitely many independent copies of the quantum state or large numbers of identical process repetitions. Instead, it focusses on finite instances of the quantum state. In this regime, we can explore how our ability to reset a quantum system (e.g., to a known pure state) is affected when accounting for the finite-size, single-instance characteristics of the system. Single-shot tasks, such as Landauer erasure, have been studied using majorization-based frameworks [314,316–319].

For example, an erasure protocol can be considered as a process involving a unital channel [320] that acts on the system whose state is to be erased (i.e., mapped to a fixed pure state), together with an information battery acting as a source of purity. In the simplest model, the information battery can be represented by an n-qubit system, where each qubit starts in a pure state and ends in a maximally mixed state. We define the work cost (in units of $k_B T \log 2$) of erasing an initial state ρ as the size of the smallest information battery that allows erasure of ρ. This is equivalent to finding the smallest integer n such that the transformation becomes possible:

$$\rho \otimes |0\rangle\langle 0|^{\otimes n} \xrightarrow[\text{channel}]{\text{unital}} |\psi\rangle\langle\psi| \otimes \left(\frac{\mathbb{1}}{2}\right)^{\otimes n}. \tag{93}$$

Recalling that Landauer bound states that $n \geq S(\rho)$, a correction to Landauer's bound in this setting is given by [317]

$$n \geq S(\rho) + \frac{V(\rho)}{2\sqrt{M(\rho)}}, \tag{94}$$

where $V(\rho) := \mathrm{tr}(\rho \log^2 \rho) - S(\rho)^2$ is the variance of surprisal and $M(\rho) := V(\rho) + [S(\rho) + 1/\ln 2]^2$. Note that $V(\rho) = 0$ for the maximally mixed state, which is typically used as the starting point for erasure. This correction indicates that the bound increases as the initial variance of ρ.

Another variant of this regime, which also allows the study of finite-size corrections in Landauer erasure, is the so-called thermodynamic distillation process. This process is defined as a thermodynamic transformation from an initial system characterized by a Hamiltonian H and prepared in a state ρ, to a target system characterized by a Hamiltonian $\tilde{H}$ and a state $\tilde{\rho}$ that is an eigenstate of $\tilde{H}$. This thermodynamic transformation can be modeled using the framework of thermal operations [253,321,322]. Consequently, corrections to Landauer erasure can be studied by asking for the optimal transformation to map N copies of maximally mixed states (with a trivial Hamiltonian) to N copies of a target state that is ϵ close to the ground state. The work cost can be quantified by appending a battery system, initially in an excited state, which is also transformed to the ground state during the transformation. The transformation error ϵ quantifies the quality of erasure. Recently, it was found that the erasure cost, valid for an arbitrary N, is given by [318]

$$W_{\mathrm{cost}} = \frac{N}{\beta}\left[\log 2 - \frac{\log(1-\epsilon)}{N}\right], \tag{95}$$

which recovers Landauer's cost of erasure for when $\epsilon = 0$.

Finally, other analyses of thermodynamics in the one-shot scenario, related to Landauer erasure, have been presented in Ref. [323] and applied to various other thermodynamic tasks [315,324–328]

IX. CONCLUSION

Maxwell's demon, once a provocative thought experiment challenging the second law of thermodynamics, has become a cornerstone of modern physics. This tutorial has traced the demon's journey from its classical origins to its quantum "reincarnations," revealing how its apparent paradoxes were resolved by recognizing information as a thermodynamic resource.

Stochastic and quantum thermodynamics have further expanded the demon's legacy, inspiring research into the fundamental limits of energy, computation, and information processing. Experimental realizations across diverse platforms demonstrate that Maxwell's demon—and related concepts such as Landauer erasure—are not merely theoretical constructs but tangible phenomena with technological implications.

No longer a paradox to be resolved, Maxwell's demon has become a powerful tool for exploring the links between thermodynamics and information. Its story highlights how thought experiments can challenge our assumptions and lead to new physics.

ACKNOWLEDGMENTS

We thank Jakub Czartowski, Jake Xuereb, Albert Rico, and Martí Perarnau-Llobet for valuable discussions and fruitful comments on the first version of the manuscript. The authors would like to thank the anonymous referee for their insightful comments and valuable suggestions, which significantly improved the clarity and presentation of this work. In particular, the suggestion to include Fig. 13 is especially appreciated. A.O.J. and J.B.B. acknowledge financial support from the Danish National Research Foundation Grant bigQ (DNRF 142) and VILLUM FONDEN through a research Grant No. (40864). R.C. acknowledges the Simons Foundation (Grant No. 1023171, RC), the Brazilian National Council for Scientific and Technological Development (CNPq, Grants No. 307295/2020-6 and No. 403181/2024-0), the Financiadora de Estudos e Projetos (Grant No. 1699/24 IIF-FINEP) and a guest professorship from the Otto Mønsted Foundation.

DATA AVAILABILITY

No data were created or analyzed in this study.

[1] A. Einstein, Über einen die erzeugung und verwandlung des lichtes betreffenden heuristischen gesichtspunkt, Ann. Phys. **322**, 132 (1905).

[2] A. Einstein, Covariance properties of the field equations of the theory of gravitation based on the generalized theory of relativity, Z. Math. Phys. **63**, 215 (1914).

[3] A. Einstein, Jahrbuch der radioaktivität und elektronik. Unter mitarbeit zahlreicher fachgenossen herausgegeben von j. stark-hannover 4. band, 1., 2. und 4. heft. verlag von s. hirzel, leipzig, Z. Elektrochem. **13**, 656 (1907).

[4] A. Einstein, Lichtgeschwindigkeit und statik des gravitationsfeldes, Ann. Phys. **343**, 355 (1912).

[5] E. Schrödinger, An undulatory theory of the mechanics of atoms and molecules, Phys. Rev. **28**, 1049 (1926).

[6] A. Einstein, B. Podolsky, and N. Rosen, Can quantum-mechanical description of physical reality be considered complete?, Phys. Rev. **47**, 777 (1935).

[7] J. S. Bell, On the Einstein Podolsky Rosen paradox, Phys. Phys. Fiz. **1**, 195 (1964).

[8] A. Aspect, J. Dalibard, and G. Roger, Experimental test of Bell's inequalities using time-varying analyzers, Phys. Rev. Lett. **49**, 1804 (1982).

[9] E. P. Wigner, *Philosophical Reflections and Syntheses* (Springer Berlin Heidelberg, New York, 1995), p. 247.

[10] N. Nurgalieva, and R. Renner, Testing quantum theory with thought experiments, Contemp. Phys. **61**, 193 (2020).

[11] A. Einstein, and P. Schilpp, *Autobiographical Notes* (Cambridge University Press, London, 1911).

[12] J. C. Maxwell, and P. Pesic, *Theory of Heat* (Courier Corporation, North Chelmsford, 2001).

[13] C. G. Knott, *Life and Scientific work of Peter Guthrie Tait* (Cambridge University Press, New York, 1911), Vol. 1.

[14] L. Szilard, Über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen, Z. Phys. **53**, 840 (1929).

[15] L. Szilard, On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings, Behav. Sci. **9**, 301 (1964).

[16] Throughout this tutorial, we will use both the inverse temperature, $\beta = 1/k_B T$ (where k_B is the Boltzmann constant), and the temperature T, depending on the context.

[17] C. E. Shannon, A symbolic analysis of relay and switching circuits, Electr. Eng. **57**, 713 (1938).

[18] C. E. Shannon, A mathematical theory of communication, Bell Labs Tech. J. **27**, 379 (1948).

[19] J. Von Neumann, First draft of a report on the EDVAC, IEEE Ann. Hist. Comput. **15**, 27 (1993).

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

[21] R. Landauer, Dissipation and noise immunity in computation and communication, Nature **335**, 779 (1988).

[22] R. Landauer, Information is physical, Phys. Today. **44**, 23 (1991).

[23] O. Penrose, *Foundations of Statistical Mechanics* (Elsevier, Amsterdam, 1970).

[24] C. H. Bennett, Logical reversibility of computation, IBM J. Res. Dev. **17**, 525 (1973).

[25] C. H. Bennett, The thermodynamics of computation—A review, Int. J. Theor. Phys. **21**, 905 (1982).

[26] U. Seifert, Stochastic thermodynamics: Principles and perspectives, Eur. Phys. J. B **64**, 423 (2008).

[27] J. Gemmer, M. Michel, and G. Mahler, *Quantum Thermodynamics: Emergence of Thermodynamic Behavior Within Composite Quantum Systems* (Springer Berlin Heidelberg, New York, 2004).

[28] R. Kosloff, Quantum thermodynamics: A dynamical viewpoint, Entropy **15**, 2100 (2013).

[29] J. Goold, M. Huber, A. Riera, L. D. Rio, and P. Skrzypczyk, The role of quantum information in thermodynamics—A topical review, J. Phys. A-Math. **49**, 143001 (2016).

[30] S. Vinjanampathy, and J. Anders, Quantum thermodynamics, Contemp. Phys. **57**, 545 (2016).

[31] R. Alicki, and R. Kosloff, in *Thermodynamics in the Quantum Regime* (Springer International Publishing, New York, 2018), p. 1.

[32] F. Binder, L. A. Correa, C. Gogolin, J. Anders, and G. Adesso, *Thermodynamics in the Quantum Regime: Fundamental Aspects and New Directions* (Springer International Publishing, New York, 2018).

[33] S. Deffner, and S. Campbell, *Quantum Thermodynamics* (Morgan and Claypool Publishers, San Rafael, 2019).

[34] S. Ciliberto, Experiments in stochastic thermodynamics: Short history and perspectives, Phys. Rev. X **7**, 021051 (2017).

[35] N. M. Myers, O. Abah, and S. Deffner, Quantum thermodynamic devices: From theoretical proposals to experimental reality, AVS Quantum Sci. **4**, 027101 (2022).

[36] M. B. Plenio, and V. Vitelli, The physics of forgetting: Landauer's erasure principle and information theory, Contemp. Phys. **42**, 25 (2001).

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

[38] E. Lutz, and S. Ciliberto, Information: From Maxwell's demon to Landauer's eraser, Phys. Today **68**, 30 (2015).

[39] H. Leff, and A. Rex, *Maxwell's Demon, Entropy, Information, Computing* (Taylor & Francis, Abingdon, 1990).

[40] E. Fermi, *Thermodynamics* (Dover Publications, Mineola, 1956).

[41] H. Callen, *Thermodynamics and an Introduction to Thermostatistics* (Wiley, Hoboken, 1985).

[42] Modern thermodynamic approaches propose a "fifth Law," related to Onsager's theorem [329,330], which

describes symmetry and reciprocity in transport coefficients for systems near local equilibrium fundamental laws.
[43] G. T. Landi, and M. Paternostro, Irreversible entropy production: From classical to quantum, Rev. Mod. Phys. **93**, 035008 (2021).
[44] M. Lax, Fluctuations from the nonequilibrium steady state, Rev. Mod. Phys. **32**, 25 (1960).
[45] R. Clausius, Über eine veränderte form des zweiten hauptsatzes der mechanischen wärmetheorie, Ann. Phys. **169**, 481 (1854).
[46] W. Thomson, XV.—On the dynamical theory of heat, with numerical results deduced from Mr Joule's equivalent of a thermal unit, and M. Regnault's observations on steam, Earth Environ. Sci. Trans. R. Soc. Edinb. **20**, 261 (1853).
[47] S. Carnot, *Reflections on the Motive Power of Fire: And Other Papers on the Second Law of Thermodynamics* (Dover, Mineola, NY, 2005).
[48] L. D. Landau, and E. M. Lifshitz, *Statistical Physics* (Addison-Wesley, Reading, MA, 1958).
[49] R. K. Pathria, *Statistical Mechanics* (Elsevier, 2016).
[50] The system does not need to be classical. We first introduce classical systems for convenience, since the next section will address quantum systems.
[51] C. Jarzynski, *Time* (Springer Basel, New York, 2012), p. 145.
[52] C. Jarzynski, Nonequilibrium equality for free energy differences, Phys. Rev. Lett. **78**, 2690 (1997).
[53] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality, Science **296**, 1832 (2002).
[54] The second law, as expressed in Eq. (10), can be directly derived using Jensen's inequality [331] applied to the Jarzynski equality. Specifically, for any convex function f and a random variable X, Jensen's inequality states that $f(\langle X\rangle) \leq \langle f(X)\rangle$.
[55] C. Jarzynski, *Time* (Springer, Basel, 2012), p. 145.
[56] G. E. Crooks, Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences, Phys. Rev. E **60**, 2721 (1999).
[57] M. Campisi, P. Hänggi, and P. Talkner, Colloquium: Quantum fluctuation relations: Foundations and applications, Rev. Mod. Phys. **83**, 771 (2011).
[58] M. Esposito, U. Harbola, and S. Mukamel, Nonequilibrium fluctuations, fluctuation theorems, and counting statistics in quantum systems, Rev. Mod. Phys. **81**, 1665 (2009).
[59] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, Thermodynamics of information, Nat. Phys. **11**, 131 (2015).
[60] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Yunger Halpern, The resource theory of informational nonequilibrium in thermodynamics, Phys. Rep. **583**, 1 (2015).
[61] M. Lostaglio, An introductory review of the resource theory approach to thermodynamics, Rep. Prog. Phys. **82**, 114001 (2019).
[62] L. I. Ponomarev, and I. Kurchatov, *The Quantum Dice* (CRC Press, Boca Raton, 2021).
[63] J. J. Sakurai, and J. Napolitano, *Modern Quantum Mechanics* (Cambridge University Press, New York, 2017).
[64] A. Peres, *Quantum Theory: Concepts and Methods* (Springer, New York, 1995).
[65] M. A. Nielsen, and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, New York, 2010).
[66] R. Bertlmann, and N. Friis, *Modern Quantum Theory: From Quantum Mechanics to Entanglement and Quantum Information* (Oxford University Press, Oxford, 2023).
[67] A. Bassi, K. Lochan, S. Satin, T. P. Singh, and H. Ulbricht, Models of wave-function collapse, underlying theories, and experimental tests, Rev. Mod. Phys. **85**, 471 (2013).
[68] A. Kossakowski, On quantum statistical mechanics of non-Hamiltonian systems, Rep. Math. Phys. **3**, 247 (1972).
[69] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, Completely positive dynamical semigroups of n-level systems, J. Math. Phys. **17**, 821 (1976).
[70] G. Lindblad, On the generators of quantum dynamical semigroups, Commun. Math. Phys. **48**, 119 (1976).
[71] J. Von Neumann, *Mathematical Foundations of Quantum Mechanics: New Edition* (Princeton University Press, Princeton, 2018), Vol. 53.
[72] von Neumann considers an ideal gas where mechanical properties follow classical mechanics, while quantum states act as informational labels. This separation of physical behavior and information allows him to use semipermeable membranes and the second law of thermodynamics to derive Eq. (17). See Ref. [332] for a rigorous formulation of this thought experiment in a purely operational way.
[73] S. Kullback, and R. A. Leibler, On information and sufficiency, Ann. Math. Stat. **22**, 79 (1951).
[74] R. Gallego, A. Riera, and J. Eisert, Thermal machines beyond the weak coupling regime, New J. Phys. **16**, 125009 (2014).
[75] P. Strasberg, G. Schaller, N. Lambert, and T. Brandes, Nonequilibrium thermodynamics in the strong coupling and non-Markovian regime based on a reaction coordinate mapping, New J. Phys. **18**, 073007 (2016).
[76] M. Perarnau-Llobet, H. Wilming, A. Riera, R. Gallego, and J. Eisert, Strong coupling corrections in quantum thermodynamics, Phys. Rev. Lett. **120**, 120602 (2018).
[77] P. Talkner, M. Campisi, and P. Hänggi, Fluctuation theorems in driven open quantum systems, J. Stat. Mech.: Theory Exp. **2009**, P02025 (2009).
[78] A. Sone, N. Yamamoto, T. Holdsworth, and P. Narang, Jarzynski-like equality of nonequilibrium information production based on quantum cross-entropy, Phys. Rev. Res. **5**, 023039 (2023).
[79] A. Delahunty, *From Bonbon to Cha-cha: Oxford Dictionary of Foreign Words and Phrases* (Oxford University Press, Oxford, 2008), p. 90
[80] J. Cresswell, *Little Oxford Dictionary of Word Origins* (Oxford University Press, Oxford, 2014), p. 146
[81] L. Boltzmann, Studien uber das gleichgewicht der lebenden kraft, Wissenschafiliche Abhandlungen **1**, 49 (1868).

[82] M. Smoluchowski, Experimentell nachweisbare, der üblichen Thermodynamik widersprechende Molekularph änomene, Pisma Mariana Smoluchowskiego **2**, 226 (1927).

[83] M. Smoluchowski, Gultigkeitsgrenzen des zweiten hauptsatzes der warmetheorie, Usp. Fiz. Nauk **93**, 724 (1967).

[84] R. P. Feynman, R. B. Leighton, M. Sands, and S. B. Treiman, The Feynman lectures on physics, Phys. Today **17**, 45 (1964).

[85] While purely mechanical devices cannot harness thermal fluctuations to perform useful work, supplying external energy and information to the system makes it possible. This principle was demonstrated with a molecule called rotaxane [333]. By adding light as an energy source, random Brownian motion can be overcome to produce a controlled movement that performs useful work. Importantly, this process does not contradict the second law of thermodynamics, as the external energy input drives the system away from equilibrium.

[86] N. Cottet, S. Jezouin, L. Bretheau, P. Campagne-Ibarcq, Q. Ficheux, J. Anders, A. Auffèves, R. Azouit, P. Rouchon, and B. Huard, Observing a quantum Maxwell demon at work, Proc. Natl. Acad. Sci. **114**, 7561 (2017).

[87] 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).

[88] J. M. Jauch, and J. G. Baron, Entropy, information and Szilard's paradox, Helv. Phys. Acta **45**, 220 (1972).

[89] P. Strasberg, *Quantum Stochastic Thermodynamics: Foundations and Selected Applications* (Oxford University Press, Oxford, 2022).

[90] W. H. Zurek, and Maxwell's demon, in *Frontiers of Nonequilibrium Statistical Physics* (Springer US, 1986), p. 151.

[91] W. H. Zurek, Eliminating ensembles from equilibrium statistical physics: Maxwell's demon, Szilard's engine, and thermodynamics via entanglement, Phys. Rep. **755**, 1 (2018).

[92] S. W. Kim, T. Sagawa, S. De Liberato, and M. Ueda, Quantum Szilard engine, Phys. Rev. Lett. **106**, 070401 (2011).

[93] In reality, the wall has a finite potential height λ_∞, ensuring during expansion that the tunneling time τ_t is much longer than the process time τ, keeping m well defined. During removal, as τ_t decreases to match τ in λ_0, the eigenstates delocalize and the partition function changes from $Z_m(l_{\rm eq}^m)$ to $Z(l_{\rm eq}^m) = \sum_{n=0}^N Z_n(l_{\rm eq}^m)$. The integral appearing in Eq. (31), for each m splits into $\int_{\lambda_\infty}^{\lambda_0} (\partial \ln Z_m(l_{\rm eq}^m)/\partial X) dX$ and $\int_{\lambda_0}^{0} (\partial \ln Z(l_{\rm eq}^m)/\partial X) dX$, where the first term vanishes in the quasistatic limit ($\tau \to \infty$).

[94] M. Plesch, O. Dahlsten, J. Goold, and V. Vedral, Comment on "quantum Szilard engine", Phys. Rev. Lett. **111**, 188901 (2013).

[95] S. W. Kim, K.-H. Kim, T. Sagawa, S. De Liberato, and M. Ueda, Kim et al. reply:, Phys. Rev. Lett. **111**, 188902 (2013).

[96] M. Plesch, O. Dahlsten, J. Goold, and V. Vedral, Maxwell's daemon: Information versus particle statistics, Sci. Rep. **4**, 6995 (2014).

[97] D. Gabor, Progress in holography, Rep. Prog. Phys **32**, 395 (1969).

[98] M. Planck, The theory of heat radiation, Entropie **144**, 164 (1900).

[99] G. Kirchhoff, in *Von Kirchhoff bis Planck: Theorie der Wärmestrahlung in Historisch-kritischer Darstellung* (1978), Vol. 131.

[100] A. Wachowski, L. Wachowski, K. Reeves, L. Fishburne, C.-A. Moss, H. Weaving, G. Foster, J. Pantoliano, and Z. Staenberg, *Matrix* (Warner Home Video, Burbank, CA, 1999).

[101] G. Boole, *An Investigation of the Laws of Thought: On Which are Founded the Mathematical Theories of Logic and Probabilities* (Walton and Maberly, London, 1854), Vol. 2.

[102] T. M. Cover, and J. A. Thomas, *Elements of Information Theory* (Wiley, Hoboken, 2001).

[103] I. Bengtsson, and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, New York, 2006).

[104] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, New York, 2013).

[105] J. Preskill, *Lecture Notes for Physics 229: Quantum Information and Computation* (CreateSpace Independent Publishing Platform, Scotts Valley, 2015).

[106] M. Tribus, and E. C. McIrvine, Energy and information, Sci. Am. **225**, 179 (1971).

[107] Brillouin proposed that acquiring information reduces a system's thermodynamic entropy, treating thermodynamic and information-theoretic entropy as equivalent and introducing the notion of negentropy [334,335]. While this contributed to the formulation of the thermodynamics of information, it has been criticized for equating observer-dependent knowledge with objective thermodynamic quantities [88]. Today, Brillouin's view is used mainly as a heuristic, not as a general physical law.

[108] H. Touchette, Equivalence and nonequivalence of ensembles: Thermodynamic, macrostate, and measure levels, J. Stat. Phys. **159**, 987 (2015).

[109] P. Strasberg, and A. Winter, First and second law of quantum thermodynamics: A consistent derivation based on a microscopic definition of entropy, PRX Quantum **2**, 030202 (2021).

[110] J. V. Neumann, Beweis des ergodensatzes und deshtheorems in der neuen mechanik, Z. Phys. **57**, 30 (1929).

[111] K. Sekimoto, *Stochastic Energetics* (Springer Berlin Heidelberg, New York, 2010).

[112] C. Van den Broeck, and M. Esposito, Ensemble and trajectory thermodynamics: A brief introduction, Physica A **418**, 6 (2015).

[113] P. Skrzypczyk, A. J. Short, and S. Popescu, Work extraction and thermodynamics for individual quantum systems, Nat. Commun. **5**, 4185 (2014).

[114] D. Safránek, J. M. Deutsch, and A. Aguirre, Quantum coarse-grained entropy and thermodynamics, Phys. Rev. A **99**, 010101 (2019).

[115] D. Safránek, J. M. Deutsch, and A. Aguirre, Quantum coarse-grained entropy and thermalization in closed systems, Phys. Rev. A **99**, 012103 (2019).

[116] D. Safránek, A. Aguirre, and J. M. Deutsch, Classical dynamical coarse-grained entropy and comparison with the quantum version, Phys. Rev. E **102**, 032106 (2020).

[117] D. Faiez, D. Safránek, J. M. Deutsch, and A. Aguirre, Typical and extreme entropies of long-lived isolated quantum systems, Phys. Rev. A **101**, 052101 (2020).

[118] J. Schindler, D. Safránek, and A. Aguirre, Quantum correlation entropy, Phys. Rev. A **102**, 052407 (2020).

[119] J. Schindler, P. Strasberg, N. Galke, A. Winter, and M. G. Jabbour, Unification of observational entropy with maximum entropy principles, ArXiv:2503.15612.

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

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

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

[123] S. Minagawa, M. H. Mohammady, K. Sakai, K. Kato, and F. Buscemi, Universal validity of the second law of information thermodynamics, Npj Quantum Inf. **11**, 18 (2025).

[124] M. Esposito, K. Lindenberg, and C. Van den Broeck, Entropy production as correlation between system and reservoir, New J. Phys. **12**, 013013 (2010).

[125] 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).

[126] T. Sagawa, and M. Ueda, Nonequilibrium thermodynamics of feedback control, Phys. Rev. E **85**, 021104 (2012).

[127] K. Micadei, R. M. Serra, and L. C. Céleri, Thermodynamic cost of acquiring information, Phys. Rev. E **88**, 062123 (2013).

[128] Y. Guryanova, N. Friis, and M. Huber, Ideal projective measurements have infinite resource costs, Quantum **4**, 222 (2020).

[129] A. Danageozian, M. M. Wilde, and F. Buscemi, Thermodynamic constraints on quantum information gain and error correction: A triple trade-off, PRX Quantum **3**, 020318 (2022).

[130] J. Xuereb, A. de Oliveira Junior, F. Clivaz, P. Bakhshinezhad, and M. Huber, What resources do agents need to acquire knowledge in quantum thermodynamics? ArXiv:2410.18167.

[131] T. E. Ouldridge, R. A. Brittain, and P. R. ten Wolde, *The Energetics of Computing in Life and Machines* (SFI Press, 2019), p. 307.

[132] 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).

[133] A. Turing, in *The Essential Turing* (Oxford University Press, Oxford, 2004), p. 58.

[134] J. Von Neumann, and A. W. Burks, *Theory of Self-Reproducing Automata* (University of Illinois Press, Urbana, 1966).

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

[136] R. Gaudenzi, E. Burzurí, S. Maegawa, H. Van Der Zant, and F. Luis, Quantum Landauer erasure with a molecular nanomagnet, Nat. Phys. **14**, 565 (2018).

[137] A. E. Allahverdyan, and T. M. Nieuwenhuizen, Breakdown of the Landauer bound for information erasure in the quantum regime, Phys. Rev. E **64**, 056117 (2001).

[138] J. D. Norton, Waiting for landauer, Stud. Hist. Philos. Sci. B: Stud. Hist. Philos. Mod. Phys. **42**, 184 (2011).

[139] R. Alicki, Quantum memory as a perpetuum mobile? stability vs. reversibility of information processing, Open Syst. Inf. Dyn. **19**, 1250016 (2012).

[140] 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, Jpn. J. Appl. Phys. **51**, 06FE10 (2012).

[141] D. Reeb, and M. M. Wolf, An improved Landauer principle with finite-size corrections, New J. Phys. **16**, 103011 (2014).

[142] L. Masanes, and J. Oppenheim, A general derivation and quantification of the third law of thermodynamics, Nat. Commun. **8**, 14538 (2017).

[143] H. Wilming, and R. Gallego, Third law of thermodynamics as a single inequality, Phys. Rev. X **7**, 041033 (2017).

[144] P. Taranto, F. Bakhshinezhad, A. Bluhm, R. Silva, N. Friis, M. P. Lock, G. Vitagliano, F. C. Binder, T. Debarba, E. Schwarzhans, F. Clivaz, and M. Huber, Landauer versus Nernst: What is the true cost of cooling a quantum system?, PRX Quantum **4**, 010332 (2023).

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

[146] J. R. Moffitt, Y. R. Chemla, S. B. Smith, and C. Bustamante, Recent advances in optical tweezers, Annu. Rev. Biochem. **77**, 205 (2008).

[147] K. Sekimoto, and S.-I. Sasa, Complementarity relation for irreversible process derived from stochastic energetics, J. Phys. Soc. Jpn. **66**, 3326 (1997).

[148] Y. Jun, M. C. V. Gavrilov, and J. Bechhoefer, High-precision test of Landauer's principle in a feedback trap, Phys. Rev. Lett. **113**, 190601 (2014).

[149] A. Bérut, A. Petrosyan, and S. Ciliberto, Information and thermodynamics: Experimental verification of Landauer's erasure principle, J. Stat. Mech.: Theory Exp. **2015**, P06015 (2015).

[150] M. C. V. Gavrilov, and J. Bechhoefer, Erasure without work in an asymmetric double-well potential, Phys. Rev. Lett. **117**, 200601 (2016).

[151] M. A. Ciampini, T. Wenzl, M. Konopik, G. Thalhammer, M. Aspelmeyer, E. Lutz, and N. Kiesel, Experimental nonequilibrium memory erasure beyond Landauer's bound, ArXiv:2107.04429.

[152] S. Dago, J. Pereda, N. Barros, S. Ciliberto, and L. Bellon, Information and thermodynamics: Fast and precise approach to Landauer's bound in an underdamped micromechanical oscillator, Phys. Rev. Lett. **126**, 170601 (2021).

[153] S. Dago, and L. Bellon, Dynamics of information erasure and extension of Landauer's bound to fast processes, Phys. Rev. Lett. **128**, 070604 (2022).

[154] J. Hong, B. Lambson, S. Dhuey, and J. Bokor, Experimental test of Landauer's principle in single-bit operations on nanomagnetic memory bits, Sci. Adv. **2**, e1501492 (2016).
[155] L. Martini, M. Pancaldi, M. Madami, P. Vavassori, G. Gubbiotti, S. Tacchi, F. Hartmann, M. Emmerling, S. Höfling, L. Worschech, and G. Carlotti, Experimental and theoretical analysis of Landauer erasure in nano-magnetic switches of different sizes, Nano Energy **19**, 108 (2016).
[156] O.-P. Saira, M. H. Matheny, R. Katti, W. Fon, G. Wimsatt, J. P. Crutchfield, S. Han, and M. L. Roukes, Nonequilibrium thermodynamics of erasure with superconducting flux logic, Phys. Rev. Res. **2**, 013249 (2020).
[157] J. P. S. Peterson, R. S. Sarthour, A. M. Souza, I. S. Oliveira, J. Goold, K. Modi, D. O. Soares-Pinto, and L. C. Céleri, Experimental demonstration of information to energy conversion in a quantum system at the Landauer limit, Proc. R. Soc. A: Math. Phys. Eng. Sci. **472**, 20150813 (2016).
[158] 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. **120**, 210601 (2018).
[159] É. Roldán, I. A. Martinez, J. M. Parrondo, and D. Petrov, Universal features in the energetics of symmetry breaking, Nat. Phys. **10**, 457 (2014).
[160] D. Barker, M. Scandi, S. Lehmann, C. Thelander, K. A. Dick, M. Perarnau-Llobet, and V. F. Maisi, Experimental verification of the work fluctuation-dissipation relation for information-to-work conversion, Phys. Rev. Lett. **128**, 040602 (2022).
[161] M. Scandi, D. Barker, S. Lehmann, K. A. Dick, V. F. Maisi, and M. Perarnau-Llobet, Minimally dissipative information erasure in a quantum dot via thermodynamic length, Phys. Rev. Lett. **129**, 270601 (2022).
[162] R. Munroe, XKCD—A webcomic of romance, sarcasm, math, and language, 2025, https://xkcd.com/ (accessed: 2025-03-06).
[163] C. H. Bennett, Time/space trade-offs for reversible computation, SIAM J. Comput. **18**, 766 (1989).
[164] D. H. Wolpert, and A. Kolchinsky, Thermodynamics of computing with circuits, New J. Phys. **22**, 063047 (2020).
[165] W. H. Zurek, Reversibility and stability of information processing systems, Phys. Rev. Lett. **53**, 391 (1984).
[166] W. H. Zurek, Thermodynamic cost of computation, algorithmic complexity and the information metric, Nature **341**, 119 (1989).
[167] E. Fredkin, and T. Toffoli, Conservative logic, Int. J. Theor. Phys **21**, 219 (1982).
[168] P. Benioff, The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by turing machines, J. Stat. Phys. **22**, 563 (1980).
[169] P. Benioff, Quantum mechanical models of turing machines that dissipate no energy, Phys. Rev. Lett. **48**, 1581 (1982).
[170] E. F. Fredkin, and T. Toffoli, *Collision-Based Computing* (Springer, London, 2002), p. 27.
[171] S. Deffner, and C. Jarzynski, Information processing and the second law of thermodynamics: An inclusive, Hamiltonian approach, Phys. Rev. X **3**, 041003 (2013).
[172] A. B. Boyd, D. Mandal, and J. P. Crutchfield, Thermodynamics of modularity: Structural costs beyond the Landauer bound, Phys. Rev. X **8**, 031036 (2018).
[173] P. Faist, and R. Renner, Fundamental work cost of quantum processes, Phys. Rev. X **8**, 021011 (2018).
[174] K. Korzekwa, and M. Lostaglio, Quantum advantage in simulating stochastic processes, Phys. Rev. X **11**, 021019 (2021).
[175] T. Conte, E. DeBenedictis, N. Ganesh, T. Hylton, J. P. Strachan, R. S. Williams, A. Alemi, L. Altenberg, G. Crooks, J. Crutchfield, *et al.*, Thermodynamic computing, ArXiv:1911.01968.
[176] D. H. Wolpert, The stochastic thermodynamics of computation, J. Phys. A-Math. **52**, 193001 (2019).
[177] T. E. Ouldridge, and D. H. Wolpert, Thermodynamics of deterministic finite automata operating locally and periodically, New J. Phys. **25**, 123013 (2023).
[178] G. Manzano, G. Kardeş, E. Roldán, and D. H. Wolpert, Thermodynamics of computations with absolute irreversibility, unidirectional transitions, and stochastic computation times, Phys. Rev. X **14**, 021026 (2024).
[179] M. T. Mitchison, Quantum thermal absorption machines: Refrigerators, engines and clocks, Contemp. Phys. **60**, 164 (2019).
[180] P. Lipka-Bartosik, M. Perarnau-Llobet, and N. Brunner, Thermodynamic computing via autonomous quantum thermal machines, Sci. Adv. **10** (2024).
[181] J. Xuereb, P. Erker, F. Meier, M. T. Mitchison, and M. Huber, Impact of imperfect timekeeping on quantum control, Phys. Rev. Lett. **131**, 160204 (2023).
[182] J. Xuereb, T. Debarba, M. Huber, and P. Erker, Quantum coding with finite thermodynamic resources, Quantum Sci. Technol. **10**, 025030 (2025).
[183] F. Meier, M. Huber, P. Erker, and J. Xuereb, Autonomous quantum processing unit: What does it take to construct a self-contained model for quantum computation? ArXiv:2402.00111.
[184] T. Sagawa, in *Thermodynamics of Information Processing in Small Systems* (Springer, Japan, 2012), p. 75.
[185] P. Strasberg, G. Schaller, T. Brandes, and M. Esposito, Thermodynamics of a physical model implementing a Maxwell demon, Phys. Rev. Lett. **110**, 040601 (2013).
[186] G. J. Milburn, An atom-optical Maxwell demon, Aust. J. Phys. **51**, 1 (1998).
[187] G. Schaller, C. Emary, G. Kiesslich, and T. Brandes, Probing the power of an electronic Maxwell's demon: Single-electron transistor monitored by a quantum point contact, Phys. Rev. B **84**, 085418 (2011).
[188] M. Esposito, and C. Van den Broeck, Second law and Landauer principle far from equilibrium, EPL **95**, 40004 (2011).
[189] D. V. Averin, M. Möttönen, and J. P. Pekola, Maxwell's demon based on a single-electron pump, Phys. Rev. B **84**, 245448 (2011).
[190] 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).
[191] A. C. Barato, and U. Seifert, An autonomous and reversible Maxwell's demon, EPL **101**, 60001 (2013).

[192] S. Deffner, Information-driven current in a quantum Maxwell demon, Phys. Rev. E **88**, 062128 (2013).

[193] 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).

[194] A. Chapman, and A. Miyake, How an autonomous quantum Maxwell demon can harness correlated information, Phys. Rev. E **92**, 062125 (2015).

[195] N. Shiraishi, S. Ito, K. Kawaguchi, and T. Sagawa, Role of measurement-feedback separation in autonomous Maxwell's demons, New J. Phys. **17**, 045012 (2015).

[196] A. Kutvonen, J. Koski, and T. Ala-Nissila, Thermodynamics and efficiency of an autonomous on-chip Maxwell's demon, Sci. Rep. **6**, 21126 (2016).

[197] K. Ptaszyński, Autonomous quantum Maxwell's demon based on two exchange-coupled quantum dots, Phys. Rev. E **97**, 012116 (2018).

[198] J. P. Pekola, D. S. Golubev, and D. V. Averin, Maxwell's demon based on a single qubit, Phys. Rev. B **93**, 024501 (2016).

[199] B.-L. Najera-Santos, P. A. Camati, V. Métillon, M. Brune, J.-M. Raimond, A. Auffèves, and I. Dotsenko, Autonomous Maxwell's demon in a cavity QED system, Phys. Rev. Res. **2**, 032025 (2020).

[200] B. Annby-Andersson, P. Samuelsson, V. F. Maisi, and P. P. Potts, Maxwell's demon in a double quantum dot with continuous charge detection, Phys. Rev. B **101**, 165404 (2020).

[201] S. Ciliberto, Autonomous out-of-equilibrium Maxwell's demon for controlling the energy fluxes produced by thermal fluctuations, Phys. Rev. E **102**, 050103 (2020).

[202] S. Seah, S. Nimmrichter, and V. Scarani, Maxwell's lesser demon: A quantum engine driven by pointer measurements, Phys. Rev. Lett. **124**, 100603 (2020).

[203] N. Freitas, and M. Esposito, Characterizing autonomous Maxwell demons, Phys. Rev. E **103**, 032118 (2021).

[204] N. Freitas, and M. Esposito, Maxwell demon that can work at macroscopic scales, Phys. Rev. Lett. **129**, 120602 (2022).

[205] S. Hernández-Gómez, S. Gherardini, N. Staudenmaier, F. Poggiali, M. Campisi, A. Trombettoni, F. Cataliotti, P. Cappellaro, and N. Fabbri, Autonomous dissipative Maxwell's demon in a diamond spin qutrit, PRX Quantum **3**, 020329 (2022).

[206] N. Freitas, and M. Esposito, Information flows in macroscopic Maxwell's demons, Phys. Rev. E **107**, 014136 (2023).

[207] R. Bao, Z. Cao, J. Zheng, and Z. Hou, Designing autonomous Maxwell's demon via stochastic resetting, Phys. Rev. Res. **5**, 043066 (2023).

[208] P. A. Erdman, R. Czupryniak, B. Bhandari, A. N. Jordan, F. Noé, J. Eisert, and G. Guarnieri, Artificially intelligent Maxwell's demon for optimal control of open quantum systems, Quantum Sci. Technol. **10**, 025047 (2025).

[209] B. Annby-Andersson, D. Bhattacharyya, P. Bakhshinezhad, D. Holst, G. De Sousa, C. Jarzynski, P. Samuelsson, and P. P. Potts, Maxwell's demon across the quantum-to-classical transition, Phys. Rev. Res. **6**, 043216 (2024).

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

[211] K. Jacobs, Second law of thermodynamics and quantum feedback control: Maxwell's demon with weak measurements, Phys. Rev. A **80**, 012322 (2009).

[212] F. J. Cao, and M. Feito, Thermodynamics of feedback controlled systems, Phys. Rev. E **79**, 041118 (2009).

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

[214] J. M. Horowitz, and J. M. R. Parrondo, Designing optimal discrete-feedback thermodynamic engines, New J. Phys. **13**, 123019 (2011).

[215] M. Esposito, and G. Schaller, Stochastic thermodynamics for "Maxwell demon" feedbacks, EPL **99**, 30003 (2012).

[216] D. Abreu, and U. Seifert, Thermodynamics of genuine nonequilibrium states under feedback control, Phys. Rev. Lett. **108**, 030601 (2012).

[217] K. Funo, Y. Watanabe, and M. Ueda, Integral quantum fluctuation theorems under measurement and feedback control, Phys. Rev. E **88**, 052121 (2013).

[218] J. P. Pekola, Towards quantum thermodynamics in electronic circuits, Nat. Phys. **11**, 118 (2015).

[219] J. V. Koski, V. F. Maisi, T. Sagawa, and J. P. Pekola, Experimental observation of the role of mutual information in the nonequilibrium dynamics of a Maxwell demon, Phys. Rev. Lett. **113**, 030601 (2014).

[220] R. Sánchez, and M. Büttiker, Optimal energy quanta to current conversion, Phys. Rev. B **83**, 085428 (2011).

[221] J. M. Horowitz, and M. Esposito, Thermodynamics with continuous information flow, Phys. Rev. X **4**, 031015 (2014).

[222] N. Shiraishi, and T. Sagawa, Fluctuation theorem for partially masked nonequilibrium dynamics, Phys. Rev. E **91**, 012130 (2015).

[223] A. Kutvonen, T. Sagawa, and T. Ala-Nissila, Thermodynamics of information exchange between two coupled quantum dots, Phys. Rev. E **93**, 032147 (2016).

[224] J. Monsel, M. Acciai, R. Sánchez, and J. Splettstoesser, Autonomous demon exploiting heat and information at the trajectory level, Phys. Rev. B **111**, 045419 (2025).

[225] H. Thierschmann, R. Sánchez, B. Sothmann, F. Arnold, C. Heyn, W. Hansen, H. Buhmann, and L. W. Molenkamp, Three-terminal energy harvester with coupled quantum dots, Nat. Nanotechnol. **10**, 854 (2015).

[226] F. Hartmann, P. Pfeffer, S. Höfling, M. Kamp, and L. Worschech, Voltage fluctuation to current converter with coulomb-coupled quantum dots, Phys. Rev. Lett. **114**, 146805 (2015).

[227] B. Roche, P. Roulleau, T. Jullien, Y. Jompol, I. Farrer, D. Ritchie, and D. Glattli, Harvesting dissipated energy with a mesoscopic ratchet, Nat. Commun. **6**, 6738 (2015).

[228] R. Sánchez, P. Samuelsson, and P. P. Potts, Autonomous conversion of information to work in quantum dots, Phys. Rev. Res. **1**, 033066 (2019).

[229] R. Sánchez, J. Splettstoesser, and R. S. Whitney, Nonequilibrium system as a demon, Phys. Rev. Lett. **123**, 216801 (2019).

[230] I. A. Picatoste, and R. Sánchez, All-thermal reversal of heat currents using qutrits, Phys. Rev. A **110**, 022210 (2024).

[231] P. A. Camati, J. P. S. Peterson, T. B. Batalhão, K. Micadei, A. M. Souza, R. S. Sarthour, I. S. Oliveira, and R. M. Serra, Experimental rectification of entropy production by Maxwell's demon in a quantum system, Phys. Rev. Lett. **117**, 240502 (2016).
[232] K. Ptaszyński, and M. Esposito, Thermodynamics of quantum information flows, Phys. Rev. Lett. **122**, 150603 (2019).
[233] M. Ribezzi-Crivellari, and F. Ritort, Large work extraction and the Landauer limit in a continuous Maxwell demon, Nat. Phys. **15**, 660 (2019).
[234] G. Manzano, D. Subero, O. Maillet, R. Fazio, J. P. Pekola, and E. Roldán, Thermodynamics of gambling demons, Phys. Rev. Lett. **126**, 080603 (2021).
[235] S. Ryu, R. López, and R. Toral, Quantum consensus dynamics by entangling Maxwell demon, New J. Phys. **24**, 033028 (2022).
[236] J. P. Garrahan, and F. Ritort, Generalized continuous Maxwell demons, Phys. Rev. E **107**, 034101 (2023).
[237] J. Oppenheim, M. Horodecki, P. Horodecki, and R. Horodecki, Thermodynamical approach to quantifying quantum correlations, Phys. Rev. Lett. **89**, 180402 (2002).
[238] W. H. Zurek, Quantum discord and Maxwell's demons, Phys. Rev. A **67**, 012320 (2003).
[239] K. Maruyama, F. Morikoshi, and V. Vedral, Thermodynamical detection of entanglement by Maxwell's demons, Phys. Rev. A **71**, 012108 (2005).
[240] K. Funo, Y. Watanabe, and M. Ueda, Thermodynamic work gain from entanglement, Phys. Rev. A **88**, 052319 (2013).
[241] K. V. Hovhannisyan, M. Perarnau-Llobet, M. Huber, and A. Acín, Entanglement generation is not necessary for optimal work extraction, Phys. Rev. Lett. **111**, 240401 (2013).
[242] H. C. Braga, C. C. Rulli, T. R. de Oliveira, and M. S. Sarandy, Maxwell's demons in multipartite quantum correlated systems, Phys. Rev. A **90**, 042338 (2014).
[243] M. Huber, M. Perarnau-Llobet, K. V. Hovhannisyan, P. Skrzypczyk, C. Klöckl, N. Brunner, and A. Acín, Thermodynamic cost of creating correlations, New J. Phys. **17**, 065008 (2015).
[244] D. E. Bruschi, M. Perarnau-Llobet, N. Friis, K. V. Hovhannisyan, and M. Huber, Thermodynamics of creating correlations: Limitations and optimal protocols, Phys. Rev. E **91**, 032118 (2015).
[245] M. Perarnau-Llobet, K. V. Hovhannisyan, M. Huber, P. Skrzypczyk, N. Brunner, and A. Acín, Extractable work from correlations, Phys. Rev. X **5**, 041011 (2015).
[246] N. Friis, M. Huber, and M. Perarnau-Llobet, Energetics of correlations in interacting systems, Phys. Rev. E **93**, 042135 (2016).
[247] A. V. Lebedev, D. Oehri, G. B. Lesovik, and G. Blatter, Trading coherence and entropy by a quantum Maxwell demon, Phys. Rev. A **94**, 052133 (2016).
[248] M. Brunelli, M. G. Genoni, M. Barbieri, and M. Paternostro, Detecting Gaussian entanglement via extractable work, Phys. Rev. A **96**, 062311 (2017).
[249] M. A. Ciampini, L. Mancino, A. Orieux, C. Vigliar, P. Mataloni, M. Paternostro, and M. Barbieri, Experimental extractable work-based multipartite separability criteria, npj Quantum Inf. **3**, 10 (2017).
[250] G. L. Zanin, M. Antesberger, M. J. Jacquet, P. H. S. Ribeiro, L. A. Rozema, and P. Walther, Enhanced photonic Maxwell's demon with correlated baths, Quantum **6**, 810 (2022).
[251] A. de Oliveira Junior, J. Son, J. Czartowski, and N. H. Y. Ng, Entanglement generation from athermality, Phys. Rev. Res. **6**, 033236 (2024).
[252] A. de Oliveira Junior, J. B. Brask, and P. Lipka-Bartosik, Heat as a witness of quantum properties, Phys. Rev. Lett. **134**, 050401 (2025).
[253] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and T. Beth, Thermodynamic cost of reliability and low temperatures: Tightening Landauer's principle and the second law, Int. J. Theor. Phys. **39**, 2717 (2000).
[254] H. E. D. Scovil, and E. O. Schulz-DuBois, Three-level masers as heat engines, Phys. Rev. Lett. **2**, 262 (1959).
[255] R. Alicki, The quantum open system as a model of the heat engine, J. Phys. A: Math. Gen. **12**, L103 (1979).
[256] R. Kosloff, A quantum mechanical open system as a model of a heat engine, J. Chem. Phys **80**, 1625 (1984).
[257] S. Lloyd, Quantum-mechanical Maxwell's demon, Phys. Rev. A **56**, 3374 (1997).
[258] T. D. Kieu, The second law, Maxwell's demon, and work derivable from quantum heat engines, Phys. Rev. Lett. **93**, 140403 (2004).
[259] T. D. Kieu, Quantum heat engines, the second law and Maxwell's daemon, Eur. Phys. J. D **39**, 115 (2006).
[260] D. Mandal, H. T. Quan, and C. Jarzynski, Maxwell's refrigerator: An exactly solvable model, Phys. Rev. Lett. **111**, 030602 (2013).
[261] A. B. Boyd, D. Mandal, and J. P. Crutchfield, Correlation-powered information engines and the thermodynamics of self-correction, Phys. Rev. E **95**, 012152 (2017).
[262] T. Fu, J. Du, S. Su, G. Su, and J. Chen, Quantum thermodynamic pump driven by Maxwell's demon, Eur. Phys. J. Plus **136**, 1059 (2021).
[263] T.-J. Zhou, Y.-H. Ma, and C. P. Sun, Finite-time optimization of a quantum Szilard heat engine, Phys. Rev. Res. **6**, 043001 (2024).
[264] H. T. Quan, Y. D. Wang, Y.-x. Liu, C. P. Sun, and F. Nori, Maxwell's demon assisted thermodynamic cycle in superconducting quantum circuits, Phys. Rev. Lett. **97**, 180402 (2006).
[265] H. T. Quan, Y.-X. Liu, C. P. Sun, and F. Nori, Quantum thermodynamic cycles and quantum heat engines, Phys. Rev. E **76**, 031105 (2007).
[266] C. Y. Cai, H. Dong, and C. P. Sun, Multiparticle quantum Szilard engine with optimal cycles assisted by a Maxwell's demon, Phys. Rev. E **85**, 031114 (2012).
[267] C. Elouard, D. Herrera-Martí, B. Huard, and A. Auffèves, Extracting work from quantum measurement in Maxwell's demon engines, Phys. Rev. Lett. **118**, 260603 (2017).
[268] K. Poulsen, M. Majland, S. Lloyd, M. Kjaergaard, and N. T. Zinner, Quantum Maxwell's demon assisted by non-Markovian effects, Phys. Rev. E **105**, 044141 (2022).
[269] F.-J. Chan, Y.-T. Huang, J.-D. Lin, H.-Y. Ku, J.-S. Chen, H.-B. Chen, and Y.-N. Chen, Maxwell's two-demon

engine under pure dephasing noise, Phys. Rev. A **106**, 052201 (2022).
[270] B. Piechocinska, Information erasure, Phys. Rev. A **61**, 062314 (2000).
[271] S. Hilt, S. Shabbir, J. Anders, and E. Lutz, Landauer's principle in the quantum regime, Phys. Rev. E **83**, 030102 (2011).
[272] B. Lambson, D. Carlton, and J. Bokor, Exploring the thermodynamic limits of computation in integrated systems: Magnetic memory, nanomagnetic logic, and the Landauer limit, Phys. Rev. Lett. **107**, 010604 (2011).
[273] L. D. Rio, J. Åberg, R. Renner, O. Dahlsten, and V. Vedral, The thermodynamic meaning of negative entropy, Nature **476**, 476 (2011).
[274] G. Diana, G. B. Bagci, and M. Esposito, Finite-time erasing of information stored in fermionic bits, Phys. Rev. E **87**, 012111 (2013).
[275] S. Deffner, and C. Jarzynski, Information processing and the second law of thermodynamics: An inclusive, Hamiltonian approach, Phys. Rev. X **3**, 041003 (2013).
[276] P. R. Zulkowski, and M. R. DeWeese, Optimal finite-time erasure of a classical bit, Phys. Rev. E **89**, 052140 (2014).
[277] S. Lorenzo, R. McCloskey, F. Ciccarello, M. Paternostro, and G. M. Palma, Landauer's principle in multipartite open quantum system dynamics, Phys. Rev. Lett. **115**, 120403 (2015).
[278] P. R. Zulkowski, and M. R. DeWeese, Optimal control of overdamped systems, Phys. Rev. E **92**, 032117 (2015).
[279] P. R. Zulkowski, and M. R. DeWeese, Optimal protocols for slowly driven quantum systems, Phys. Rev. E **92**, 032113 (2015).
[280] C. Browne, A. J. P. Garner, O. C. O. Dahlsten, and V. Vedral, Guaranteed energy-efficient bit reset in finite time, Phys. Rev. Lett. **113**, 100603 (2014).
[281] J. Goold, M. Paternostro, and K. Modi, Nonequilibrium quantum Landauer principle, Phys. Rev. Lett. **114**, 060602 (2015).
[282] P. Strasberg, G. Schaller, T. Brandes, and M. Esposito, Quantum and information thermodynamics: A unifying framework based on repeated interactions, Phys. Rev. X **7**, 021003 (2017).
[283] G. Guarnieri, S. Campbell, J. Goold, S. Pigeon, B. Vacchini, and M. Paternostro, Full counting statistics approach to the quantum non-equilibrium Landauer bound, New J. Phys. **19**, 103038 (2017).
[284] S. Campbell, G. Guarnieri, M. Paternostro, and B. Vacchini, Nonequilibrium quantum bounds to Landauer's principle: Tightness and effectiveness, Phys. Rev. A **96**, 042109 (2017).
[285] J. Klaers, Landauer's erasure principle in a squeezed thermal memory, Phys. Rev. Lett. **122**, 040602 (2019).
[286] A. M. Timpanaro, J. P. Santos, and G. T. Landi, Landauer's principle at zero temperature, Phys. Rev. Lett. **124**, 240601 (2020).
[287] N. Shiraishi, and K. Saito, Information-theoretical bound of the irreversibility in thermal relaxation processes, Phys. Rev. Lett. **123**, 110603 (2019).
[288] M. Scandi, and M. Perarnau-Llobet, Thermodynamic length in open quantum systems, Quantum **3**, 197 (2019).
[289] K. Proesmans, J. Ehrich, and J. Bechhoefer, Finite-time Landauer principle, Phys. Rev. Lett. **125**, 100602 (2020).
[290] H. J. D. Miller, G. Guarnieri, M. T. Mitchison, and J. Goold, Quantum fluctuations hinder finite-time information erasure near the Landauer limit, Phys. Rev. Lett. **125**, 160602 (2020).
[291] K. Proesmans, J. Ehrich, and J. Bechhoefer, Optimal finite-time bit erasure under full control, Phys. Rev. E **102**, 032105 (2020).
[292] Y.-Z. Zhen, D. Egloff, K. Modi, and O. Dahlsten, Universal bound on energy cost of bit reset in finite time, Phys. Rev. Lett. **127**, 190602 (2021).
[293] T. Van Vu, and Y. Hasegawa, Lower bound on irreversibility in thermal relaxation of open quantum systems, Phys. Rev. Lett. **127**, 190601 (2021).
[294] T. Van Vu, and K. Saito, Finite-time quantum Landauer principle and quantum coherence, Phys. Rev. Lett. **128**, 010602 (2022).
[295] Y.-H. Ma, J.-F. Chen, C. P. Sun, and H. Dong, Minimal energy cost to initialize a bit with tolerable error, Phys. Rev. E **106**, 034112 (2022).
[296] Y.-Z. Zhen, D. Egloff, K. Modi, and O. Dahlsten, Inverse linear versus exponential scaling of work penalty in finite-time bit reset, Phys. Rev. E **105**, 044147 (2022).
[297] A. B. Boyd, A. Patra, C. Jarzynski, and J. P. Crutchfield, Shortcuts to thermodynamic computing: The cost of fast and faithful information processing, J. Stat. Phys. **187**, 17 (2022).
[298] J. S. Lee, S. Lee, H. Kwon, and H. Park, Speed limit for a highly irreversible process and tight finite-time Landauer's bound, Phys. Rev. Lett. **129**, 120603 (2022).
[299] P. Chattopadhyay, A. Misra, T. Pandit, and G. Paul, Landauer principle and thermodynamics of computation, ArXiv:quantum-ph/2506.10876.
[300] J. Czartowski, A. de Oliveira Junior, and K. Korzekwa, Thermal recall: Memory-assisted Markovian thermal processes, PRX Quantum **4**, 040304 (2023).
[301] A. de Oliveira Junior, Geometric and information-theoretic aspects of quantum thermodynamics, ArXiv: 2404.00617.
[302] P. Abiuso, H. J. D. Miller, M. Perarnau-Llobet, and M. Scandi, Geometric optimisation of quantum thermodynamic processes, Entropy **22**, 1076 (2020).
[303] Finite-time processes in quantum thermodynamics: The limits of irreversibility, PhD thesis, 2024, https://archive-ouverte.unige.ch/unige:180515.
[304] A. Rolandi, M. Perarnau-Llobet, and H. J. D. Miller, Optimal control of dissipation and work fluctuations for rapidly driven systems, New J. Phys. **25**, 073005 (2023).
[305] A. Rolandi, and M. Perarnau-Llobet, Finite-time Landauer principle beyond weak coupling, Quantum **7**, 1161 (2023).
[306] S. A. Hartnoll, and A. P. Mackenzie, Colloquium: Planckian dissipation in metals, Rev. Mod. Phys. **94**, 041002 (2022).
[307] N. W. Ashcroft, N. D. Mermin, and R. Smoluchowski, Solid state physics, Phys. Today **30**, 61 (1977).
[308] P. Lipka-Bartosik, and M. Perarnau-Llobet, Minimizing dissipation via interacting environments: Quadratic convergence to Landauer bound, ArXiv:2411.00944.
[309] D. Reeb, and M. M. Wolf, An improved Landauer principle with finite-size corrections, New J. Phys. **16**, 103011 (2014).

[310] E. Bäumer, M. Perarnau-Llobet, P. Kammerlander, H. Wilming, and R. Renner, Imperfect thermalizations allow for optimal thermodynamic processes, Quantum **3**, 153 (2019).

[311] P. Taranto, P. Lipka-Bartosik, N. A. Rodríguez-Briones, M. Perarnau-Llobet, N. Friis, M. Huber, and P. Bakhshinezhad, Efficiently cooling quantum systems with finite resources: Insights from thermodynamic geometry, Phys. Rev. Lett. **134**, 070401 (2025).

[312] M. Tomamichel, *Quantum Information Processing with Finite Resources* (Springer International Publishing, New York, 2016).

[313] D. Egloff, O. C. O. Dahlsten, R. Renner, and V. Vedral, A measure of majorization emerging from single-shot statistical mechanics, New J. Phys. **17**, 073001 (2015).

[314] G. Gour, Quantum resource theories in the single-shot regime, Phys. Rev. A **95**, 062314 (2017).

[315] P. Lipka-Bartosik, and R. Demkowicz-Dobrzański, Thermodynamic work cost of quantum estimation protocols, J. Phys. A-Math. **51**, 474001 (2018).

[316] C. T. Chubb, M. Tomamichel, and K. Korzekwa, Beyond the thermodynamic limit: finite-size corrections to state interconversion rates, Quantum **2**, 108 (2018).

[317] P. Boes, N. H. Ng, and H. Wilming, Variance of relative surprisal as single-shot quantifier, PRX Quantum **3**, 010325 (2022).

[318] T. Biswas, A. de Oliveira Junior, M. Horodecki, and K. Korzekwa, Fluctuation-dissipation relations for thermodynamic distillation processes, Phys. Rev. E **105**, 054127 (2022).

[319] A. de Oliveira Junior, J. Czartowski, K. Życzkowski, and K. Korzekwa, Geometric structure of thermal cones, Phys. Rev. E **106**, 064109 (2022).

[320] Those are channels that preserve the maximally mixed state.

[321] M. Horodecki, and J. Oppenheim, Fundamental limitations for quantum and nanoscale thermodynamics, Nat. Commun. **4**, 2059 (2013).

[322] F. G. S. L. Brandão, M. Horodecki, N. H. Y. Ng, J. Oppenheim, and S. Wehner, The second laws of quantum thermodynamics, Proc. Natl. Acad. Sci. USA **112**, 3275 (2015).

[323] N. Y. Halpern, A. J. P. Garner, O. C. O. Dahlsten, and V. Vedral, Introducing one-shot work into fluctuation relations, New J. Phys. **17**, 095003 (2015).

[324] M. Lostaglio, M. P. Müller, and M. Pastena, Stochastic independence as a resource in small-scale thermodynamics, Phys. Rev. Lett. **115**, 150402 (2015).

[325] J. Gemmer, and J. Anders, From single-shot towards general work extraction in a quantum thermodynamic framework, New J. Phys. **17**, 085006 (2015).

[326] J. G. Richens, and L. Masanes, Work extraction from quantum systems with bounded fluctuations in work, Nat. Commun. **7**, 13511 (2016).

[327] S. Salek, and K. Wiesner, Fluctuations in single-shot ϵ-deterministic work extraction, Phys. Rev. A **96**, 052114 (2017).

[328] S.-Y. Wang, Single-shot work extraction in quantum thermodynamics revisited, J. Phys. A-Math. **51**, 035305 (2017).

[329] L. Onsager, Reciprocal relations in irreversible processes. I, Phys. Rev. **37**, 405 (1931).

[330] L. Onsager, Reciprocal relations in irreversible processes. II., Phys. Rev. **38**, 2265 (1931).

[331] J. L. W. V. Jensen, Sur les fonctions convexes et les inégalités entre les valeurs moyennes, Acta Math. **30**, 175 (1906).

[332] S. Minagawa, H. Arai, and F. Buscemi, Von Neumann's information engine without the spectral theorem, Phys. Rev. Res. **4**, 033091 (2022).

[333] V. Serreli, C.-F. Lee, E. R. Kay, and D. A. Leigh, Nature **445**, 523 (2007).

[334] J. Rothstein, Information, measurement, and quantum mechanics, Science **114**, 171 (1951).

[335] L. Brillouin, The negentropy principle of information, J. Appl. Phys. **24**, 1152 (1953).