

# Refining Landauer’s Stack: Balancing Error and Dissipation When Erasing Information

Gregory W. Wimsatt,<sup>1,\*</sup> Alexander B. Boyd,<sup>2,†</sup> Paul M. Riechers,<sup>2,‡</sup> and James P. Crutchfield<sup>1,§</sup>

<sup>1</sup>*Complexity Sciences Center and Physics Department,  
University of California at Davis, One Shields Avenue, Davis, CA 95616*

<sup>2</sup>*Complexity Institute and School of Physical and Mathematical Sciences,  
Nanyang Technological University, 637371 Singapore, Singapore*

(Dated: February 14, 2021)

Nonequilibrium information thermodynamics determines the minimum energy dissipation to reliably erase memory under time-symmetric control protocols. We demonstrate that its bounds are tight and so show that the costs overwhelm those implied by Landauer’s energy bound on information erasure. Moreover, in the limit of perfect computation, the costs diverge. The conclusion is that time-asymmetric protocols should be developed for efficient, accurate thermodynamic computing. And, that Landauer’s Stack—the full suite of theoretically-predicted thermodynamic costs—is ready for experimental test and calibration.

Keywords: nonequilibrium steady state, thermodynamics, dissipation, entropy production, Landauer bound

## I. INTRODUCTION

In 1961, Landauer identified a fundamental energetic requirement to perform logically-irreversible computations on nonvolatile memory [1]. Focusing on arguably the simplest case—erasing a bit of information—he found that one must supply at least  $k_B T \ln 2$  work energy ( $\approx 10^{-21} J$  at room temperature), eventually expelling this as heat. (Here,  $k_B$  is Boltzmann’s constant and  $T$  is the temperature of the computation’s ambient environment.) Notably, though still underappreciated, Landauer had identified a thermodynamically-reversible transformation. And so, no entropy actually need be produced—energy is not irrevocably dissipated—at least in the quasistatic, thermodynamically-reversible limit required to meet Landauer’s bound.

Landauer’s original argument appealed to equilibrium statistical mechanics. Since his time, advances in nonequilibrium thermodynamics, though, showed that his bound on the required work follows from a modern version of the Second Law of thermodynamics [2]. (And, when the physical substrate’s dynamics are taken into account, this is the *information processing Second Law* (IPSL) [3].) These modern laws clarified many connections between information processing and thermodynamics, such as dissipation bounds due to system-state coarse-grainings [4], nanoscale information-heat engines [5], the relation of dissipation and fluctuating currents [6], and memory design [7].

Additional scalings recently emerged between computation time, space, reliability, thermodynamic efficiency, and robustness of information storage [8–10]. In contrast to Landauer’s bound, these tradeoffs involve thermodynamically-irreversible processes, implying that entropy production and therefore true heat dissipation is generally required depending on either practicality or design goals.

In addition to these tradeoffs, it is now clear that substantial energetic costs are incurred when using logic gates and allied information-processing modules to construct a computer. Especially so, when compared to custom designing hardware to optimally implement a particular computation [11].

Taken altogether these costs constitute a veritable *Landauer’s Stack* of the information-energy requirements for thermodynamic computing. Figure 1 illustrates Landauer’s Stack in the light of historical trends in the thermodynamic costs of performing elementary logic operations in CMOS technology. The units there are joules dissipated per logic operation. We take Landauer’s Stack to be the overhead including Landauer’s bound ( $k_B T \ln 2$  joules) up to the current (year 2020) energy dissipations *due to information processing*. Thus, the Stack is a hierarchy of energy expenditures that underlie contemporary digital computing—an arena of theoretically-predicted and as-yet unknown thermodynamic phenomena waiting detailed experimental exploration.

To account for spontaneous deviations that arise in small-scale systems, the Second Laws are now most properly expressed by exact equalities on probability distributions of possible energy fluctuations. These are the *fluctuation theorems* [24], from which the original Laws (in fact, inequalities) can be readily recovered. Augmenting the

---

\* gwwimsatt@ucdavis.edu

† abboyd@ucdavis.edu

‡ pmriechers@gmail.com

§ chaos@ucdavis.edu

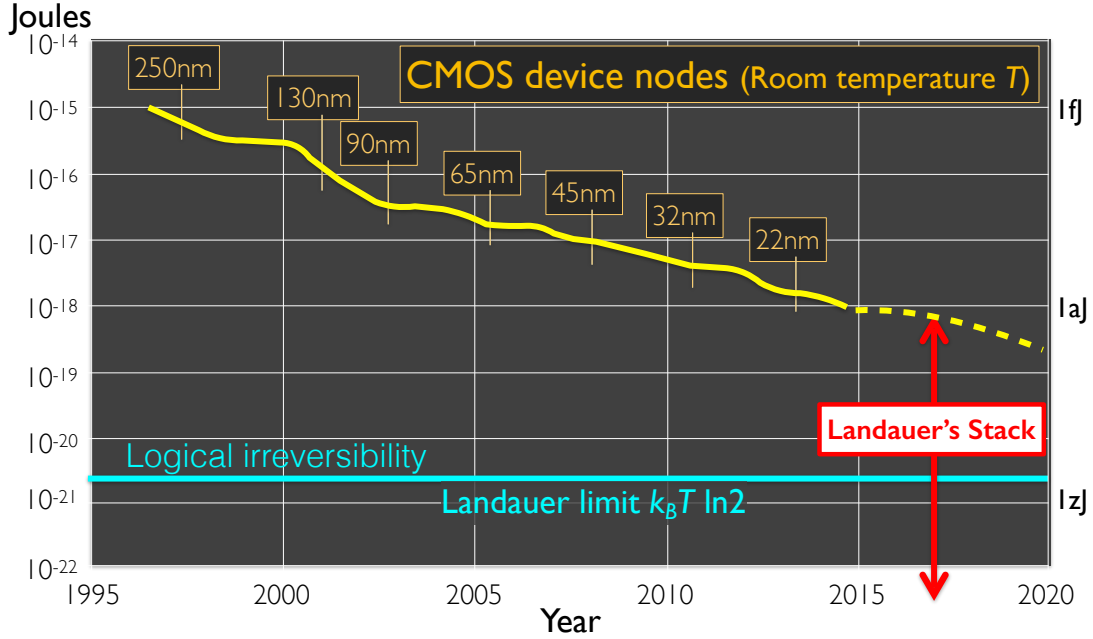


FIG. 1. Historical trends in thermodynamic costs of performing elementary logic operations in CMOS technology quoted in energy dissipated (joules) per logic operation. Contemporary experimentally-accessible thermal resolution is approximately  $10^{-24}$  joules. Landauer’s Stack—Thermodynamic hierarchy of predicted “overhead” energy expenditures *due to information processing* that underlie contemporary digital computing, including Landauer’s Principle of logical irreversibility [1] (which is now seen as a consequence of the broader information processing Second Law  $\langle W \rangle \leq k_B T \ln \Delta h_\mu$  [3]): (a) Nonreciprocity [12]; (b) Computation rate  $1/\tau$  [9, 10]; (c) Accuracy:  $-\ln \epsilon$  [12]; (d) Storage stability; (e) Circuit modularity [11]; (f) Mismatched expectations [13, 14]; (g) Transitions between nonequilibrium steady-state storage states [15, 16]; and (h) Quantum coherence [17]. (2015 and prior portion of figure courtesy M. L. Roukes, data compiled from [18, and citations therein]. Landauer’s Stack cf. Table I in Ref. [19].) CMOS technology change to 3D device nodes around 2015 make linear feature size and its relation to energy costs largely incomparable afterwards [20–23]. There are, of course, other sources of energy dissipation in CMOS such as leakage currents that arise when electrons tunnel from gate to drain through a thin gate-oxide dielectric. Thermodynamically, this source is a kind of “housekeeping heat”, necessary to support the substrate’s electronic properties but not directly due to information processing.

Stack, fluctuation theorems apply directly to information processing, elucidating further thermodynamic restrictions and structure [25–28].

The result is a rather more complete accounting for the energetic costs of thermodynamic computation, captured in the refined Landauer’s Stack of Fig. 1. In this spirit, here we report new bounds on the work required to compute in the very important case of computations driven externally by time-symmetric control protocols [12]. In surprising contrast to the fixed energy cost of erasure identified by Landauer, here we demonstrate that the scaling of the minimum required energy *diverges as a function of accuracy* and so can dominate Landauer’s Stack. This serves the main goal in the following to validate and demonstrate the tightness of Ref. [12]’s thermodynamic bounds and do so in Landauer’s original setting of information erasure.

In essence, our argument is as follows. Energy dissipation in thermodynamic transformations is strongly related to entropy production. The fluctuation theorems establish that entropy production depends on both forward and

reverse dynamics. Thus, when determining bounds on dissipation in thermodynamic computing, one has to examine both when the control protocol is applied in forward and reverse. By considering time-symmetric protocols we substantially augment Landauer and Bennett’s dissipation bound on logical irreversibility [29] with dissipation due to logical nonselfinvertibility (aka nonreciprocity). Our results therefore complement recent work on the consequences of logical and thermodynamic reversibility [30]. Parallel work on thermodynamic bounds for information processing in finite time, and bit-erasure in particular, include the use of optimized control in the linear response regime [31–33] and transport theory [34–36]. However, the cost of nonreciprocity necessarily goes beyond the cost of finite-time computing, because time-symmetrically driven computations incur this additional dissipation regardless of the rate at which they’re executed.

Why time-symmetric protocols? Modern digital computers are driven by sinusoidal line voltages and square-wave clock pulses. These control signals function as control

parameters, directly altering the energetics and therefore guiding dynamics of the computer components. Being time-symmetric control signals, modern digital computers must then obey Ref. [12]’s error-dissipation trade-off. Moreover, the costs apply to even the most basic of computational tasks—such as bit erasure. Here, we present protocols for time-symmetrically implementing erasure in two different frameworks and demonstrate that both satisfy the new bounds. Moreover, many protocols approach the bounds quite closely, indicating that they may in fact be broadly achievable.

After a brief review of the general theory, we begin with an analysis of erasure implemented with the simple framework of two-state rate equations, demonstrating the validity of the bound for different protocols of increasing reliability. We then expand our framework to fully simulated collections of particles erased in an underdamped Langevin double-well potential, seeing the same faithfulness to the bound for a wide variety of different erasure protocols. We conclude with a call for follow-on efforts to analyze even more efficient computing that can arise from *time-asymmetric* protocols.

## II. DISSIPATION IN THERMODYNAMIC COMPUTING

Consider a universe consisting of a computing device—the *system under study* (SUS), a *thermal environment* at fixed inverse temperature  $\beta$ , and a *laboratory device* (lab) that includes a *work reservoir*. The set of possible microstates for the SUS is denoted  $\mathcal{S}$ , with  $s$  denoting an individual SUS microstate. The SUS is driven by a *control parameter*  $x$  generated by the lab. The SUS is also in contact with the thermal environment.

The overall evolution occurs from time  $t = 0$  to  $t = \tau$  and is determined by two components. The first is the SUS’s Hamiltonian  $\mathcal{H}_{SL}(s, x)$  that specifies its interaction with the lab device and determines (part of) the rates of change of the SUS coordinates consistent with Hamiltonian mechanics. We refer to the possible values of the Hamiltonian as the *SUS energies*. The second component is the thermal environment which exerts a stochastic influence on the system dynamics.

We design the lab to guarantee that a specific control parameter value  $x(t)$  is applied to the SUS at every time  $t$  over the time interval  $t \in (0, \tau)$ . That is, the control parameter evolves deterministically as a function of time. The deterministic trajectory taken by the control parameter  $x(t)$  over the computation interval is the *control protocol*, denoted by  $\vec{x}$ . The SUS microstate  $s(t)$  exhibits a response to the control protocol over the interval, following a stochastic trajectory denoted  $\vec{s}$ .

For a given microstate trajectory  $\vec{s}$ , the net energy transferred from the lab to the SUS is defined as the *work*, which has the following form [5]:

$$W(\vec{s}) = \int_0^\tau dt \dot{x}(t) \frac{\partial \mathcal{H}_{SL}}{\partial x} \Big|_{(s(t), x(t))}.$$

This is the energy accumulated in the SUS directly caused by changes in the control parameter.

Given an initial microstate  $s_0$ , the probability of a microstate trajectory  $\vec{s}$  conditioned on starting in  $s_0$  is denoted:

$$P(\vec{s}|s_0) = \Pr_{\vec{x}}(\vec{\mathcal{S}} = \vec{s} | \mathcal{S}_0 = s_0).$$

With the SUS initialized in microstate distribution  $\mu_0$ , the unconditioned *forward process* gives the probability of trajectory  $\vec{s}$ :

$$P(\vec{s}) = P(\vec{s}|s(0))\mu_0(s(0)).$$

*Detailed fluctuation theorems* (DFTs) [37, 38] determine thermodynamic properties of the computation by comparing the forward process to the *reverse process*. This requires determining the conditional probability of trajectories under time-reversed control:

$$R(\vec{s}|s_0) = \Pr_{\vec{\mathbf{x}}}(\vec{\mathcal{S}} = \vec{s} | \mathcal{S}_0 = s_0).$$

The reverse control protocol is  $\mathbf{J}\mathbf{x}(t) = x(\tau - t)^\dagger$ , where  $x^\dagger$  is  $x$ , but with all time-odd components (e.g., magnetic field) flipped in sign. And, the *reverse process* results from the application of this dynamic to the final distribution  $\mu_\tau$  of the forward process with microstates conjugated:

$$R(\vec{s}) = R(\vec{s}|s(0))\mu_\tau(s(0)^\dagger).$$

The Crooks DFT [38] then gives an equality on both the dissipated work (or entropy production) that is produced as well as the required work for a given trajectory induced by the protocol:

$$\beta W_{\text{diss}}(\vec{s}) = \ln \frac{P(\vec{s})}{R(\mathbf{J}\vec{s})}.$$

$\mathbf{J}\vec{s}$  here is itself a SUS microstate trajectory with  $\mathbf{J}s(t) = s(\tau - t)^\dagger$ .

Due to their practical relevance, we consider protocols that are symmetric under time reversal  $\mathbf{J}\vec{x} = \vec{x}$ . That is, the reverse-process probability of trajectory  $\vec{s}$  conditioned on starting in microstate  $s_0$  is the same as that of the forward process:

$$R(\vec{s}|s_0) = P(\vec{s}|s_0).$$

However, the unconditional reverse-process probability of the trajectory  $\vec{s}$  is then:

$$R(\vec{s}) = P(\vec{s}|s(0))\mu_\tau(s(0)^\dagger) .$$

This leads to a version of Crooks' DFT that can be used to set modified bounds on a computation's dissipation:

$$\beta W_{\text{diss}}(\vec{s}) = \ln \frac{P(\vec{s}|s(0))\mu_0(s(0))}{P(\mathbf{J}\vec{s}|s(\tau)^\dagger)\mu_\tau(s(\tau))} . \quad (1)$$

Suppose, now, that the final and initial SUS Hamiltonian configurations  $\mathcal{H}_{SL}(s, x(\tau))$  and  $\mathcal{H}_{SL}(s, x(0))$  are both designed to store the same information about the SUS. The SUS microstates are partitioned into locally-stable regions that are separated by large energy barriers in these energy landscapes. On some time scale, a state initialized in one of these regions has a very low probability of escape and instead locally equilibrates to its locally-stable region. These regions can thus be used to store information for periods of time controlled by the energy barrier heights. Collectively, we refer to these regions as *memory states*  $\mathcal{M}$ .

Then the probability of the system evolving to a memory state  $m' \in \mathcal{M}$  given that it starts in a memory state  $m \in \mathcal{M}$  under either the forward or reverse process is:

$$P'(m \rightarrow m') = \frac{\int d\vec{s} [s(0) \in m \wedge s(\tau) \in m'] P(\vec{s})}{\int d\vec{s} [s(0) \in m] P(\vec{s})} ,$$

where  $[E]$  evaluates to one if expression  $E$  is true and zero otherwise.

To simplify the development, suppose that the energy landscape of each memory state looks the same locally. That is, up to translation and possibly reflection and rotation, each memory state spans the same volume in microstate space and has the same energies at each of those states. Further, suppose that the SUS starts and ends in a metastable distribution, differing from global equilibrium only in the weight that each memory state is given in the distribution. Otherwise the distribution looks identical to the global equilibrium at the local scale of any memory state. This ensures that the average change in SUS energy is zero, simplifying the change  $\Delta$  in nonequilibrium free energy  $\mathcal{F}$  [12]:

$$\Delta\mathcal{F} = -\beta^{-1}\Delta H(\mathcal{M}_t) ,$$

where  $H(\cdot)$  is the Shannon entropy (in nats), and  $\mathcal{M}_t$  is the random variable for the memory state at time  $t$ . Finally, suppose that the time reversal of a microstate changes neither the memory state it exists in nor its equilibrium probability, for any time during the protocol. This holds for memory states distinguished primarily by the

positions of the system particles and system Hamiltonians that are unchanging under time reversal. See Ref. [12] for details behind these assumptions and generalized bounds without them.

Then we have the following inequality:

$$\beta\langle W_{\text{diss}} \rangle \geq \Delta H(\mathcal{M}_t) + \sum_{m \in \mathcal{M}} \mu'_0(m) \sum_{m' \in \mathcal{M}} d(m, m'), \quad (2)$$

where:

$$\begin{aligned} \mu'_0(m) &= \int ds [s \in m] \mu_0(s) \quad \text{and} \\ d(m, m') &= P'(m \rightarrow m') \ln \frac{P'(m \rightarrow m')}{P'(m' \rightarrow m)} . \end{aligned}$$

See Appendix A for a proof sketch.

Recalling that  $\beta\langle W_{\text{diss}}(\vec{s}) \rangle = \beta(\langle W(\vec{s}) \rangle - \Delta\mathcal{F})$  and appealing to the inequality in Eq. (2), we find a simple bound on the average work over the protocol:

$$\begin{aligned} \beta\langle W \rangle &\geq \sum_{m \in \mathcal{M}} \mu'_0(m) \sum_{m' \in \mathcal{M}} d(m, m') \quad (3) \\ &\equiv \beta\langle W \rangle_{\min}^{\text{t-sym}} . \end{aligned}$$

This provides a bound on the work that depends solely on the logical operation of the computation, but goes beyond Landauer's bound.

Since we are addressing modern computing, we consider processes that approximate deterministic computations on the memory states. For such computations there exists a computation function  $C : \mathcal{M} \rightarrow \mathcal{M}$  such that the physically-implemented stochastic map approximates the desired function up to some small error. That is,  $P'(m \rightarrow C(m)) = 1 - \epsilon_m$  where  $0 < \epsilon_m \ll 1$ . In fact, we require all relevant errors to be bound by a small error-threshold  $\epsilon \ll 1$ . That is, for all  $C(m) \neq m'$ , let  $P'(m, m') = \epsilon_{m \rightarrow m'}$  so that  $0 \leq \sum_{m' \neq C(m)} \epsilon_{m \rightarrow m'} = \epsilon_m \leq \epsilon \ll 1$ .

We can then simplify Eq. (3)'s bound in the limit of small  $\epsilon$ . First, we show that  $d(m, m') \geq 0$  for any pair of  $m, m'$  in the small  $\epsilon$  limit, where we have:

$$\begin{aligned} d(m, m') &= P'(m \rightarrow m') \ln \frac{P'(m \rightarrow m')}{P'(m' \rightarrow m)} \\ &\geq P'(m \rightarrow m') \ln P'(m \rightarrow m') . \end{aligned}$$

If  $C(m) = m'$ , then  $P'(m \rightarrow m') = 1 - \epsilon_m \geq 1 - \epsilon$ , so that:

$$d(m, m') \geq (1 - \epsilon) \ln(1 - \epsilon) ,$$

which vanishes as  $\epsilon \rightarrow 0$ . And, if  $C(m) \neq m'$ , then

$P'(m \rightarrow m') = \epsilon_{m \rightarrow m'}$ , so that:

$$d(m, m') \geq \epsilon_{m \rightarrow m'} \ln \epsilon_{m \rightarrow m'} ,$$

which also vanishes as  $\epsilon \rightarrow 0$ . Setting this asymptotic lower bound on the dissipation of each transition facilitates isolating divergent contributions, such as those we now consider.

An *unreciprocated* memory transition  $C(m) = m'$  is one that does not map back to itself:  $C(m') \neq m$ . The contribution to the dissipation bound is:

$$\begin{aligned} d(m, m') &= (1 - \epsilon_m) \ln \frac{1 - \epsilon_m}{\epsilon_{m' \rightarrow m}} \\ &\geq (1 - \epsilon) \ln \frac{1 - \epsilon}{\epsilon} . \end{aligned}$$

As  $\epsilon \rightarrow 0$ , this gives:

$$d(m, m') \geq \ln \epsilon^{-1} . \quad (4)$$

That is, as computational accuracy increases ( $\epsilon \rightarrow 0$ ),  $d(m, m')$  diverges. This means the minimum-required work (Eq. (3)) must then also diverge.

We then arrive at our simplified bound for the small- $\epsilon$  high-accuracy limit from Eq. (3)'s inequality on dissipation by only including the contribution from unreciprocated transitions  $m' = C(m)$  for which  $m \neq C(m')$ :

$$\begin{aligned} \beta \langle W \rangle &\geq \ln(\epsilon^{-1}) \sum_{m \in \mathcal{M}} \mu'_0(m) [C(C(m)) \neq m] \quad (5) \\ &\equiv \beta \langle W \rangle_{\min}^{\text{approx}} . \end{aligned}$$

In this way, we see how computational accuracy drives a thermodynamic cost that diverges, overwhelming the Landauer-erasure cost. A similar logarithmic relationship between dissipated work and error was demonstrated in the context of the adaptation accuracy of *Escherichia coli* and other simple biological systems [39].

The bound in Eq. (5) also applies to digital computing such as that performed with dynamic random-access memory (DRAM). We recognize that its operation places the device in a nonequilibrium steady state, appearing to negate the applicability of Crooks' fluctuation theorem in Eq. (1). However, the remedy for systems whose steady states are nonequilibrium is simply to replace the equality with an inequality, implying that more work must be dissipated than in the case of a local-equilibrium steady state [40]. Thus, our derived bounds must still hold for these modern computing devices.

### III. ERASURE THERMODYNAMICS

Inequalities Eqs. (3) and (5) place severe constraints on the work required to process information via time-symmetric control on memories. The question remains, though, whether or not these bounds can actually be met by specific protocols or if there might be still tighter bounds to be discovered.

To help answer this question, we turn to the case, originally highlighted by Landauer [1], of erasing a single bit of information. This remarkably simple case of computing has held disproportionate sway in the development of thermodynamic computing compared to other elementary operations. The following does not deviate from this habit, showing, in fact, that there remain fundamental issues. We explore this via two different implementations: The first, described via two-state rate equations, and the second with an underdamped double-well potential—Landauer's original, preferred setting.

Suppose the SUS supports two (mesoscopic) memory states, labeled L and R. The task of a time-symmetric protocol that implements erasure is to guide the SUS microscopic dynamics that starts with an initial 50 – 50 distribution over the two memory states to a final distribution as biased as possible onto the L state. The logical function  $C$  of perfect bit erasure is attained when  $C(L) = C(R) = L$ , setting either memory state to L. The probabilities of incorrectly sending an L state to R and an R state to R are denoted  $\epsilon_L$  and  $\epsilon_R$ , respectively.

Error generation is described by the binary asymmetric channel [41]—the *erasure channel*  $\mathcal{E}$  with conditional probabilities:

$$\mathcal{E} = \begin{matrix} & \mathcal{M}_\tau = L & \mathcal{M}_\tau = R \\ \begin{matrix} \mathcal{M}_0 = L \\ \mathcal{M}_0 = R \end{matrix} & \begin{pmatrix} 1 - \epsilon_L & \epsilon_L \\ 1 - \epsilon_R & \epsilon_R \end{pmatrix} \end{matrix} .$$

For any erasure implementation, this Markov transition matrix gives the error rate  $\epsilon_L = \epsilon_{L \rightarrow R}$  from initial memory state  $\mathcal{M}_0 = L$  and the error rate  $\epsilon_R = \epsilon_{R \rightarrow R}$  from the initial memory state  $\mathcal{M}_0 = R$ .

Noting first that  $d(m, m) = 0$  generically, we then have:

$$\begin{aligned} d(L, R) &= \epsilon_L \ln \frac{\epsilon_L}{1 - \epsilon_R} , \\ d(R, L) &= (1 - \epsilon_R) \ln \frac{1 - \epsilon_R}{\epsilon_L} . \end{aligned}$$

So, the bound of Eq. (3) simplifies to:

$$\begin{aligned} \beta \langle W \rangle_{\min}^{\text{t-sym}} &= \frac{1}{2} \epsilon_L \ln \frac{\epsilon_L}{1 - \epsilon_R} + \frac{1}{2} (1 - \epsilon_R) \ln \frac{1 - \epsilon_R}{\epsilon_L} \\ &= \left( \frac{1}{2} - \langle \epsilon \rangle \right) \ln \frac{1 - \epsilon_R}{\epsilon_L} , \quad (6) \end{aligned}$$

where  $\langle \epsilon \rangle = (\epsilon_L + \epsilon_R)/2$  is the average error for the process. Notice further that  $C(C(L)) = L$  but  $C(C(R)) \neq R$ , indicating that only the computation on R is nonreciprocal. Therefore, the bound of Eq. (5) simplifies to:

$$\beta \langle W \rangle_{\min}^{\text{approx}} = \frac{1}{2} \ln(\epsilon^{-1}) . \quad (7)$$

Applying Eq. (7) to DRAM directly provides a quantitative comparison beyond a formal divergence of energy costs. Contemporary DRAM exhibits a range of “soft” error rates around  $10^{-22}$  failures per write operation [42]. In fact, each write operation is effectively an erasure. (The quoted statistic is an average of 4,000 correctable errors per 128 MB DIMM per year.) Using Eq. (7), this gives a thermodynamic cost of  $25 k_B T$ , which is markedly larger than Landauer’s  $k_B T \ln 2$  factor. It is also, just as clearly, smaller by a factor of roughly 10 than the contemporary energy costs per logic operation displayed in Fig. 1. These numerical results on the ability to meet our bounds for the case of bit erasure support the conclusion that modern computers can still be improved in efficiency, despite that efficiency being ultimately limited by the bounds we introduced. The conclusion is further reinforced by the numerical simulations in the following sections that nearly achieve our theoretical bounds.

### A. Erasure with Two-state Rate Equations

A direct test of time-symmetric erasure requires only a simple two-state system that evolves under a rate equation:

$$\begin{aligned} & \frac{d \Pr(\mathcal{M}_t = m)}{dt} \\ &= \sum_{m'} [r_{m' \rightarrow m}(t) \Pr(\mathcal{M}_t = m') - r_{m \rightarrow m'}(t) \Pr(\mathcal{M}_t = m)], \end{aligned} \quad (8)$$

obeying the Arrhenius equations:

$$\begin{aligned} r_{R \rightarrow L}(t) &= A e^{-\Delta E_R(t)/k_B T} \text{ and} \\ r_{L \rightarrow R}(t) &= A e^{-\Delta E_L(t)/k_B T} , \end{aligned}$$

where the states are labeled  $\{L, R\}$  and the terms  $\Delta E_R(t)$  and  $\Delta E_L(t)$  in the exponentials are the activation energies to transit over the energy barrier at time  $t$  for the Right and Left wells, respectively.

These dynamics can be interpreted as a coarse-graining of thermal motion in a double-well potential energy landscape  $V(q, t)$  over the positional variable  $q$  at time  $t$ . Above,  $A$  is an arbitrary constant, which is fixed for the dynamics.  $q_R^*$  and  $q_L^*$  are the locations of the Right

and Left potential well minima, respectively. Thus, assuming that  $q = 0$  is the location of the barrier’s maximum between them, we see that the activation energies can be expressed as  $\Delta E_R(t) = V(0, t) - V(q_R^*, t)$  and  $\Delta E_L(t) = V(0, t) - V(q_L^*, t)$ . By varying the potential energy extrema  $V(q_R^*, t)$ ,  $V(q_L^*, t)$ , and  $V(0, t)$  we control the dynamics of the observed variables  $\{L, R\}$  in much the same way as is done with physical implementations of erasure where barrier height and tilt are controlled in a double-well [43].

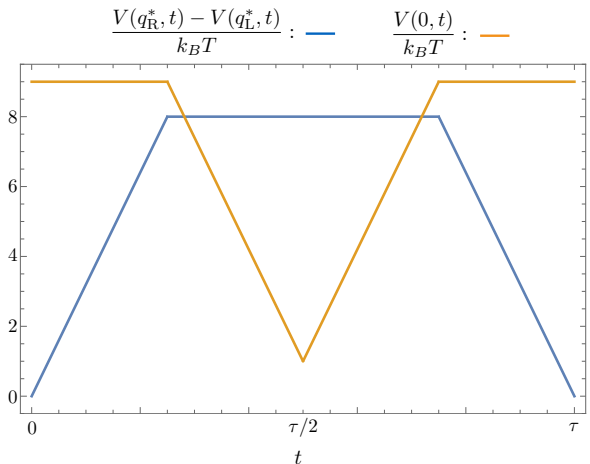


FIG. 2. Time-symmetric control protocol for implementing moderately-efficient erasure. This should be compared to Landauer’s original time-symmetric protocol [1]. It starts by tilting—increasing the difference in potential energy  $(V(q_R^*, t) - V(q_L^*, t))/k_B T$  between L and R. We increase this value such that transitions are more likely to go from R to L. Then we reduce the barrier height  $V(0, t)$  to increase the total flow rate. Finally, we reverse the previous steps, cutting off the flow by raising the barrier, then untilting.

Deviating from previous investigations of efficient erasure, where Landauer’s bound was nearly achieved over long times [43, 44], here the constraint to time-symmetric driving over the interval  $t \in (0, \tau)$  results in additional dissipated work. As Landauer described [1], erasure can be implemented by turning on and off a tilt from R to L—a time-symmetric protocol. However, to achieve higher accuracy, we also lower the barrier while the system is tilted energetically towards the L well.

Consider a family of control protocols that fit the profile shown in Fig. 2. First, we increase the energy tilt from R to L via the energy difference  $V(q_R^*, t) - V(q_L^*, t)$  measured in units of  $k_B T$ . This increases the relative probability of transitioning R to L. However, with the energy barrier at its maximum height, the transition takes quite some time. Thus, we reduce the energy barrier  $V(0, t)$  to its minimum height halfway through the protocol  $t = \tau/2$ . Then, we reverse the protocol, raising the barrier back to its default height to hold the probability distribution

fixed in the well and untilt so that the system resets to its default double-well potential.

Increasing the maximum tilt—given by  $V(q_R^*, \tau/2) - V(q_L^*, \tau/2)$  at the halfway time—increases erasure accuracy. Figure 3 shows that the maximum error  $\epsilon = \max\{\epsilon_R, \epsilon_L\}$  decreases nearly exponentially with increased maximum energy difference between left and right, going below 1 error in every 1000 trials for our parameter range. Note that  $\epsilon$  starts at a very high value (greater than  $1/2$ ) for zero tilt, since the probability  $\epsilon_R = \epsilon$  of ending in the R well starting in the R well is very high if there is no tilt to push the system out of the R well.

Figure 3 also shows the relationship between the work and the bounds described above. Given that our system consists of two states  $\{L, R\}$  and that we choose a control protocol that keeps the energy  $V(q_L^*, t)$  on the left fixed, the work (marked by green +s in the figure) is [5]:

$$\begin{aligned} \langle W \rangle &= \int_0^\tau dt \sum_s \Pr(\mathcal{S}_t = s) \partial_t V(s, t) \\ &= \int_0^\tau dt \Pr(\mathcal{M}_t = R) \partial_t V(q_R^*, t). \end{aligned}$$

This work increases almost linearly as the error reduces exponentially.

As a first comparison, note that the Landauer bound  $\langle W \rangle_{\min}^{\text{Landauer}} = -k_B T \Delta H(\mathcal{M}_t)$  (marked by orange x's in the figure) is still valid. However, it is a very weak bound for this time-symmetric protocol. The Landauer bound saturates at  $k_B T \ln 2$ . Thus, the dissipated work—the gap between orange x's and green +s—grows approximately linearly with increasing tilt energy.

In contrast, Eq. (6)'s bound  $\langle W \rangle_{\min}^{t\text{-sym}}$  for time-symmetric protocols is much tighter. The time-symmetric bound is valid: marked by blue circles that all fall below the calculated work (green +s). Not only is this bound much stricter, but it almost exactly matches the calculated work for a large range of parameters, with the work only diverging for higher tilts and lower error rates.

Finally, the approximate bound  $\langle W \rangle_{\min}^{\text{approx}} = \frac{k_B T}{2} \ln \epsilon^{-1}$  (marked by red +s) of Eq. (7), which captures the error scaling, behaves as expected. The error-dependent work bound nearly exactly matches the exact bound for low error rates on the right side of the plot and effectively bounds the work. For lower tilts, this quantity does not bound the work and is not a good estimate of the true bound, but this is consistent with expectations for high error rates. This approximation should only be employed for very reliable computations, for which it appears to be an excellent estimate. Thus, the two-level model of erasure demonstrates that the time-symmetric control bounds on work and dissipation are reasonable in both their exact and approximate forms at low error rates.

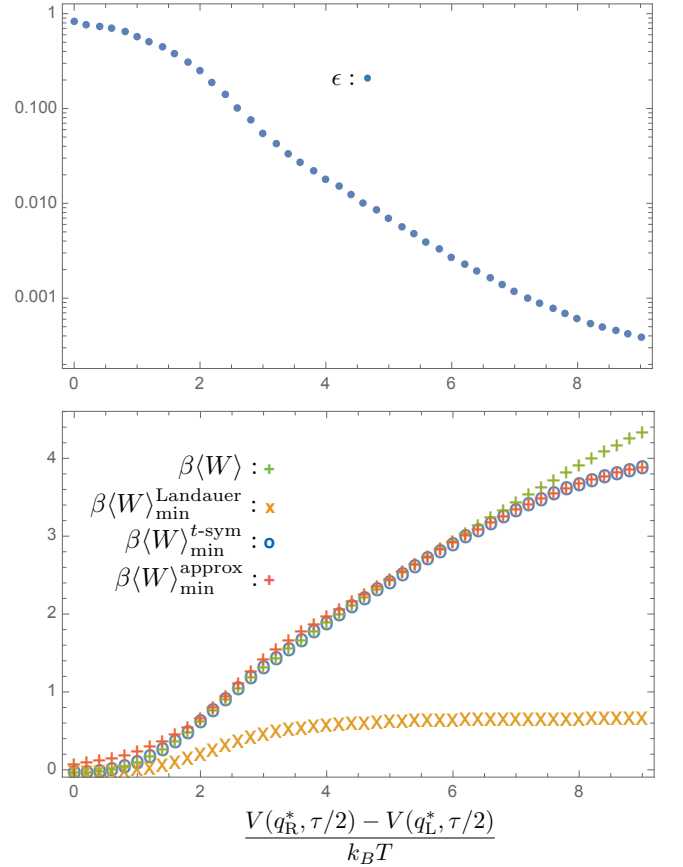


FIG. 3. (Top) Maximum error  $\epsilon$  (blue dots) decreases approximately exponentially with increasing maximum tilt. The latter is given by the maximum energy difference between the right and left energy well  $V(q_R^*, \tau/2) - V(q_L^*, \tau/2)$ . (Bottom) Work  $\langle W \rangle$  (green +s), scaled by the inverse temperature  $\beta = 1/k_B T$ , increases with increasing maximum tilt and decreasing error. The Landauer work bound  $\langle W \rangle_{\min}^{\text{Landauer}}$  (orange x's) is a very weak bound, asymptoting to a constant value rather than continuing to increase, as the work does. The bound  $\langle W \rangle_{\min}^{t\text{-sym}}$  (blue circles) on time-symmetrically driven protocols, on the other hand, is a very tight bound for lower values of maximum tilt. The work deviates from the time-symmetric bound for higher tilts. Finally, the approximate bound  $\langle W \rangle_{\min}^{\text{approx}}$  (red +s), which scales as  $\ln \epsilon^{-1}$ , is not an accurate bound over the entire range, but it very closely matches the exact time-symmetric bound  $\langle W \rangle_{\min}^{t\text{-sym}}$  for small  $\epsilon$ , as expected.

## B. Erasure with an Underdamped Double-well Potential

The physics in the rate equations above represents a simple model of a bistable thermodynamic system, which can serve as an approximation for many different bistable systems. One possible interpretation is a coarse-graining of the Langevin dynamics of a particle moving in a double-well potential. To explore the broader validity of the error-dissipation tradeoff, here we simulate the dynamics

of a stochastic particle coupled to a thermal environment at constant temperature and a work reservoir via such a 1D potential. Again, we find that the time-symmetric bounds are much tighter than Landauer's, reflecting the error-dissipation tradeoff of this control protocol class.

Consider a one-dimensional particle with position and momentum in an external potential and in thermal contact with the environment at temperature  $T$ . We consider a protocol architecture similar to that of Sec. III A, but with additional passive substages at the beginning middle and end: (i) hold the potential in the symmetric double-well form, (ii) positively tilt the potential, (iii) completely drop the potential barrier between the two wells, (iv) hold the potential while it is tilted with no barrier, (v) restore the original barrier, (vi) remove the positive tilt, restoring the original symmetric double-well, and (vii) hold the potential in this original form.

As a function of position  $q$  and time  $t$ , the potential then takes the form:

$$V(q, t) = aq^4 - b_0 b_f(t) q^2 + c_0 c_f(t) q ,$$

with constants  $a, b_0, c_0 > 0$ . The protocol functions  $b_f(t)$  and  $c_f(t)$  evolve in a piecewise linear and time-symmetric manner according to Table I, where  $t_0, t_1, \dots, t_7 = 0, \tau/12, 3\tau/12, 5\tau/12, 7\tau/12, 9\tau/12, 11\tau/12, \tau$ . The potential thus begins and ends in a symmetric double-well configuration with each well defining a memory state. During the protocol, though, the number of metastable regions is temporarily reduced to one. Figure 4 (top three panels) shows the protocol functions over time as well as the resultant potential function at key times for one such set of protocol parameters; see nondimensionalization in App. II. At any time, we label the metastable regions from most negative position to most positive the L state and, if it exists, the R state.

$t$	$t_0$	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$	$t_7$
$b_f(t)$	1	1	$\frac{t_3-t}{t_3-t_2}$	0	$\frac{t-t_4}{t_5-t_4}$	1	1	1
$c_f(t)$	0	$\frac{t-t_1}{t_2-t_1}$	1	1	1	$\frac{t_6-t}{t_6-t_5}$	0	0

TABLE I. Erasure protocol.

We simulate the motion of the particle with underdamped Langevin dynamics:

$$dq = v dt$$

$$m dv = - \left( \frac{\partial}{\partial q} V(q, t) + \lambda v \right) dt + \sqrt{2k_B T \lambda} r(t) \sqrt{dt} ,$$

where  $\lambda$  is the coupling between the thermal environment

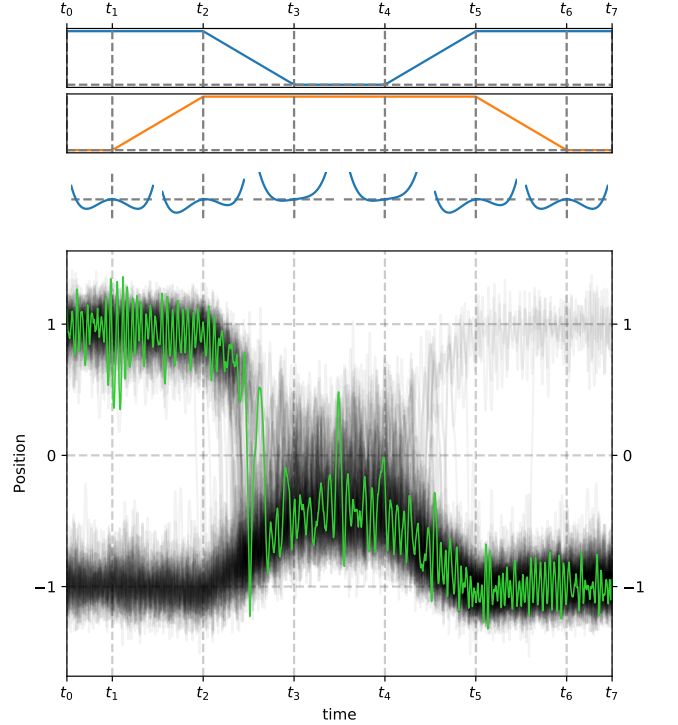


FIG. 4. Erasure via an underdamped double-well potential: Protocol functions  $b(t)$  (top panel, blue) and  $c(t)$  (second panel, orange) are symmetric in time, guaranteeing the potential function (third panel) to evolve symmetrically in time. Due to the spatial asymmetry in the potential over the majority of the protocol, however, erasure to state L ( $x < 0$ ) typically occurs, evidenced by the evolution of the system position for 100 randomly-chosen trajectories (bottom panel, black). The L and R states merge into one between times  $t_2$  and  $t_3$  and separate again between times  $t_4$  and  $t_5$ . A single trajectory (bottom panel, green) shows the typical behavior of falling into the  $x < 0$  region by time  $t_3$  and remaining there when the L state is reintroduced for the rest of the protocol.

and particle,  $m$  is the particle's mass, and  $r(t)$  is a memoryless Gaussian random variable with  $\langle r(t) \rangle = 0$  and  $\langle r(t)r(t') \rangle = \delta(t - t')$ . The particle is initialized to be in global equilibrium over the initial potential  $V(\cdot, 0)$ . Figure 4 (bottom panel) shows 100 randomly-chosen resultant trajectories for a choice of process parameters.

The work done on a single particle over the course of the protocol with trajectory  $(q(t))_t$  is [5]:

$$W = \int_0^\tau dt \frac{\partial V(q, t)}{\partial t} \Big|_{q=q(t)} .$$

Figure 5 shows the net average work over time for an erasure process, comparing it to (i) the Landauer bound, (ii) the exact bound of Eq. (6), and (iii) the approximate bound of Eq. (7). Notice that the final net average work lies above all three, as it should and that the time-symmetric bounds presented here are tighter than

Landauer's.

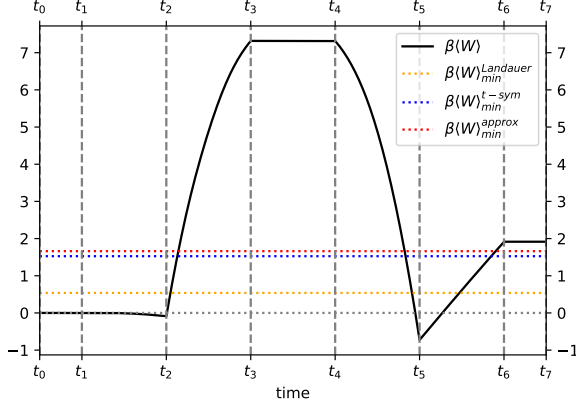


FIG. 5. Average work in  $k_B T$  over time for an erasure (black). Calculated from the simulation-estimated values  $\epsilon_L$  and  $\epsilon_R$ , Landauer's bound is given by the dashed yellow line and our approximate and exact bounds (Eqs. (7) and (6)) are given in dashed red and blue lines, respectively.

We repeat this comparison for an array of different parameters for the erasure protocol. As described in App. II, we vary features of the dynamics—including mass  $m$ , temperature  $T$ , coupling to the heat bath  $\lambda$ , duration of control  $\tau$ , maximum depth of the potential energy wells, and maximum tilt between the wells. Nondimensionalization reduces the relevant parameters to just four, allowing us to explore a broad swathe of possible physical erasures with 735 different protocols. For each protocol, we simulate 100,000 trajectories to estimate the work cost and errors  $\epsilon_R$  and  $\epsilon_L$  of the operation.

Figure 6 compares the work spent for each of the 735 erasure protocols to the sampled maximum error  $\epsilon = \max(\epsilon_L, \epsilon_R)$ . Each protocol corresponds to a green cross, whose vertical position corresponds to the shifted work  $\langle W \rangle_{\text{shift}}$ , which accounts for inhomogeneities in the error rate. Note that the exact bound  $\langle W \rangle_{\text{min}}^{t\text{-sym}}$  from Eq. (6) reduces to a simple relationship between work and error tolerance  $\epsilon$  when the errors are homogeneous  $\epsilon_R = \epsilon_L = \epsilon$ :

$$\langle W \rangle_{\text{ref}}^{t\text{-sym}} = \left( \frac{1}{2} - \epsilon \right) \ln \frac{1 - \epsilon}{\epsilon},$$

which we plot with the blue curve in Fig. 6. The cost of inhomogeneities in the error is evaluated by the difference between this reference bound and the exact work bound. This cost is added to the calculated work for each protocol to determine the shifted work:

$$\langle W \rangle_{\text{shift}} = \langle W \rangle + \langle W \rangle_{\text{ref}}^{t\text{-sym}} - \langle W \rangle_{\text{min}}^{t\text{-sym}},$$

such that the vertical distance between  $\langle W \rangle_{\text{shift}}$  and  $\langle W \rangle_{\text{ref}}^{t\text{-sym}}$  in Fig. 6 gives the true difference  $\langle W \rangle - \langle W \rangle_{\text{min}}^{t\text{-sym}}$

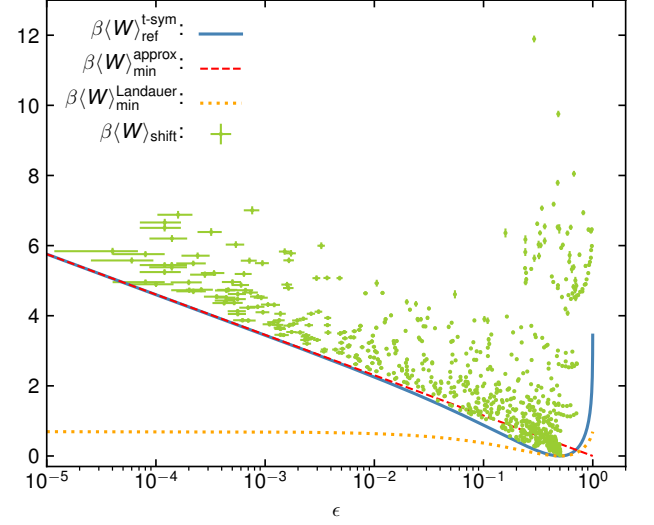


FIG. 6. Reference bound  $\langle W \rangle_{\text{ref}}^{t\text{-sym}}$  (blue line) lower bounds all of the shifted works  $\langle W \rangle_{\text{shift}}$  (green markers), often quite tightly. The approximate bound  $\langle W \rangle_{\text{min}}^{\text{approx}}$  (red dashed line) rapidly converges with decreasing error to  $\langle W \rangle_{\text{ref}}^{t\text{-sym}}$ . Time-asymmetric protocols can do better, needing only to satisfy Landauer's bound  $\langle W \rangle_{\text{min}}^{\text{Landauer}}$  (orange dotted line).

between the average sampled work and exact bound for the simulated protocol.

Figure 6 shows that the shifted average works for all of the simulated protocols in green, including error bars, all lay above the reference work bound in blue. Thus, we see that all simulated protocols satisfy the bound  $\langle W \rangle \geq \langle W \rangle_{\text{min}}^{t\text{-sym}}$ . Furthermore, many simulated protocols end up quite close to their exact bound. There are protocols with small errors, but they have larger average works. The error–dissipation tradeoff is clear.

The error–dissipation tradeoff is further illustrated in Fig. 6 by the red line, which describes the low- $\epsilon$  asymptotic bound  $\langle W \rangle_{\text{min}}^{\text{approx}}$  given by Eq. (7). In this semi-log plot, it rather quickly becomes an accurate approximation for small error.

Finally, Fig. 6 plots the Landauer bound  $\langle W \rangle_{\text{min}}^{\text{Landauer}}$  as a dotted orange line. It is calculated using the final probability of the R mesostate. The bound is weaker than that set by  $\langle W \rangle_{\text{ref}}^{t\text{-sym}}$ . As  $\epsilon \rightarrow 0$ , the gap between  $\langle W \rangle_{\text{ref}}^{t\text{-sym}}$  and  $\langle W \rangle_{\text{min}}^{\text{Landauer}}$  in Fig. 6 relentlessly increases. The stark difference in the energy scale of the time-symmetric bounds developed here and that of the looser Landauer bound shows a marked tightening of thermodynamic bounds on computation.

Notably, the protocol Landauer originally proposed to erase a bit requires *significantly more work* than his bound  $k_B T \ln 2$  to reliably erase a bit. This extra cost is a direct consequence of his protocol's time symmetry. It

turns out that time-*asymmetric* protocols for bit erasure have been used in experiments that more nearly approach Landauer’s bound [45, 46]. Although, it is not clear to what extent time asymmetry was an intentional design constraint in their construction, since there was no general theoretical guidance until now for why time-symmetry or asymmetry should matter. Figures 6 and 3 confirm that Ref. [46]’s time-asymmetric protocol for bit erasure—where the barrier is lowered before the tilt, but then raised before untilting—is capable of reliable erasure that is more thermodynamically efficient than any time-symmetric protocol could ever be.

These underdamped simulations drive home the point that our bounds are independent of the details of the dynamics used for computation. Our results are very general in that regard. As long as the system *starts* metastable and is then driven by a time-symmetric protocol, the error–dissipation tradeoff quantifies the minimal dissipation that will be incurred (for a desired level of computational accuracy) by the time the system relaxes again to metastability.

#### IV. CONCLUSION

We adapted Ref. [12]’s thermodynamic analysis of time-symmetric protocols to give a detailed analysis of the trade-offs between accuracy and dissipation encountered in erasing information.

Reference [12] showed that time symmetry and metastability together imply a generic error–dissipation tradeoff. The minimal work expected for a computation  $\mathcal{C}$  is the average nonreciprocity. In the low-error limit—where the probability of error must be much less than unity ( $\epsilon \ll 1$ )—the minimum work diverges according to:

$$\beta \langle W \rangle_{\min}^{\text{approx}} = \langle [\mathcal{C}(\mathcal{M}_0) \neq \mathcal{M}_0] \rangle_{\mathcal{M}_0} \ln(\epsilon^{-1})$$

Of all of this work, only the meager Landauer cost  $\Delta H(\mathcal{M}_t)$ , which saturates to some finite value as  $\epsilon \rightarrow 0$ , can be thermodynamically recovered in principle. Thus, irretrievable dissipation scales as  $\ln(\epsilon^{-1})$ . The reciprocity coefficient  $\langle [\mathcal{C}(\mathcal{M}_0) \neq \mathcal{M}_0] \rangle_{\mathcal{M}_0}$  depends only on the deterministic computation to be approximated. This points out likely energetic inefficiencies in current instantiations of reliable computation. It also suggests that time-asymmetric control may allow more efficient computation—but only when time-asymmetry is a free resource, in contrast to modern computer architecture.

The results here verified these general conclusions for erasure, showing in detail how tight the bounds can be and, for high-reliability thermodynamic computing, how they overwhelm Landauer’s. It may be fruitful to explore

the ideas behind our results in explicitly quantum, finite, and even zero-temperature systems. Refined versions of Landauer’s bound and other thermodynamic results can be obtained for such models [47, 48]. Also, explicit consideration of finite-time protocols can reveal efficiency advantages when treating ensembles of systems under majority-logic decoding [49–51]. Perhaps analogous refinements of the results presented here can be found as well.

Despite the almost universal focus on information erasure as a proxy for all of computing, we now see that there is a wide diversity of costs in thermodynamic computing. Looking to the future, these costs must be explored in detail if we are to design and build more capable and energy efficient computing devices. Beyond engineering and sustainability concerns, explicating Landauer’s Stack will go a long way to understanding the fundamental physics of computation—one of Landauer’s primary goals [52]. In this way, we now better appreciate the suite of thermodynamic costs—what we called Landauer’s Stack—that underlies modern computing.

#### ACKNOWLEDGMENTS

The authors thank the Telluride Science Research Center for hospitality during visits and the participants of the Information Engines Workshops there for helpful discussions. JPC acknowledges the kind hospitality of the Santa Fe Institute, Institute for Advanced Study at the University of Amsterdam, and California Institute of Technology for their hospitality during visits. This material is based upon work supported by, or in part by, FQXi Grant number FQXi-RFP-IPW-1902, Templeton World Charity Foundation Power of Information fellowship TWCF0337, and U.S. Army Research Laboratory and the U.S. Army Research Office under contracts W911NF-13-1-0390 and W911NF-18-1-0028.

## Appendix A: Proof of Exact Bound for Time-Symmetric Protocols

Here, we prove Eq. (2). This is an exact bound under the following conditions: the protocol is time-symmetric; it operates on systems that start and end in metastable equilibrium; the corresponding metastable regions look the same in state space versus energy up to translation, rotation, and reflection in state space; the memory states are symmetric under time reversal; and the equilibrium probability of a microstate at any time does not change under time reversal of the microstate. These conditions are all met in a wide variety of computational processes, including the bit erasure processes studied here. For a more general bound that applies under fewer restrictions, see Ref. [12].

We start with Eq. (1), Crooks' DFT applied to time-symmetric protocols:

$$\begin{aligned}\beta W_{\text{diss}}(\vec{s}) &= \ln \frac{P(\vec{s}|s(0))\mu_0(s(0))}{P(\mathbf{J}\vec{s}|s(\tau)^\dagger)\mu_\tau(s(\tau))} \\ &= \ln \frac{P(\vec{s})}{R(\mathbf{J}\vec{s})}.\end{aligned}$$

To help with the algebra, we introduce a second way to define reverse process probabilities:

$$Q(\vec{\mathcal{S}} = \vec{s}) = Q(\vec{s}) = R(\mathbf{J}\vec{s}) = P(\mathbf{J}\vec{s}|s(\tau)^\dagger)\mu_\tau(s(\tau)).$$

We also allow both  $P$  and  $Q$  to take as arguments system microstate trajectories  $\vec{s}$ , pairs of initial and final memory states  $\vec{m} = (m(0), m(\tau))$ , or a combination of the two, where  $\vec{m}$  may even appear as a condition. We define the time reverse of the pair of initial and final memory states as  $\mathbf{J}\vec{m} = \mathbf{J}(m(0), m(\tau)) = (m(\tau)^\dagger, m(0)^\dagger) = (m(\tau), m(0))$ . (Note that  $m^\dagger \equiv \{s^\dagger : s \in m\}$  and we explicitly assume in the following that the each memory state is time-reversal invariant; i.e.,  $m^\dagger = m$  for all  $m \in \mathcal{M}$ .) For example, the probability of observing the reverse trajectory  $\mathbf{J}\vec{s}$  in the reverse process conditioned on observing the reverse of the pair of initial and final memory states  $\mathbf{J}\vec{m}$  is:

$$\begin{aligned}Q(\vec{s}|\vec{m}) &= Q(\vec{\mathcal{S}} = \vec{s}|\vec{M} = \vec{m}) = \frac{Q(\vec{\mathcal{S}} = \vec{s}, \vec{M} = \vec{m})}{Q(\vec{M} = \vec{m})} = \frac{Q(\vec{s}, \vec{m})}{Q(\vec{m})} \\ &= \frac{\Pr_{\mathbf{J}\vec{x}}(\vec{\mathcal{S}} = \mathbf{J}\vec{s}, \vec{M} = \mathbf{J}\vec{m}|\mathcal{S}_0 = s(\tau)^\dagger)\mu_\tau(s(\tau))}{\int d\vec{s}' \Pr_{\mathbf{J}\vec{x}}(\vec{\mathcal{S}} = \mathbf{J}\vec{s}', \vec{M} = \mathbf{J}\vec{m}|\mathcal{S}_0 = s'(\tau)^\dagger)\mu_\tau(s'(\tau))},\end{aligned}$$

where  $\vec{M}$  is the random variable for the pair of initial and final memory states.

To derive Eq. (2), we first show that the average dissipated work is bounded by a Kullback–Leibler divergence on forward process memory state distributions compared with reverse process memory state distributions. Letting  $\vec{\mathcal{M}}$  be the set of all possible pairs of initial and final memory states:

$$\begin{aligned}\langle \beta W_{\text{diss}} \rangle &= \int d\vec{s} P(\vec{s}) \ln \frac{P(\vec{s})}{Q(\vec{s})} \\ &= \sum_{\vec{m} \in \vec{\mathcal{M}}} \int_{\vec{m}} d\vec{s} P(\vec{s}, \vec{m}) \ln \frac{P(\vec{s}, \vec{m})}{Q(\vec{s}, \vec{m})} \\ &= \sum_{\vec{m} \in \vec{\mathcal{M}}} \int_{\vec{m}} d\vec{s} P(\vec{s}|\vec{m}) P(\vec{m}) \ln \frac{P(\vec{s}|\vec{m}) P(\vec{m})}{Q(\vec{s}|\vec{m}) Q(\vec{m})} \\ &= \sum_{\vec{m} \in \vec{\mathcal{M}}} P(\vec{m}) \ln \frac{P(\vec{m})}{Q(\vec{m})} + \sum_{\vec{m} \in \vec{\mathcal{M}}} P(\vec{m}) \int_{\vec{m}} d\vec{s} P(\vec{s}|\vec{m}) \frac{P(\vec{s}|\vec{m})}{Q(\vec{s}|\vec{m})} \\ &= D_{\text{KL}}[P \parallel Q]_{\vec{\mathcal{M}}} + \sum_{\vec{m} \in \vec{\mathcal{M}}} P(\vec{m}) D_{\text{KL}}[P(\cdot|\vec{m}) \parallel Q(\cdot|\vec{m})]_{\vec{m}},\end{aligned}$$

where  $D_{\text{KL}}[\cdot \parallel \cdot]_{\vec{\mathcal{M}}}$  is the Kullback–Leibler divergence for distributions over all  $\vec{m} \in \vec{\mathcal{M}}$  and  $D_{\text{KL}}[\cdot \parallel \cdot]_{\vec{m}}$  is that for all

$\vec{s}$  consistent with  $\vec{m}$ . Since Kullback–Leibler divergences are nonnegative, we arrive at the following inequality:

$$\langle \beta W_{\text{diss}} \rangle \geq D_{\text{KL}} [P \parallel Q]_{\vec{\mathcal{M}}} .$$

Second, we show that the above Kullback–Leibler divergence equals the righthand side of Eq. (2).

To help with this second task, we start by establishing that  $Q(\mathcal{M}_0 = m | \mathcal{M}_\tau = m') = P(\mathcal{M}_\tau = m | \mathcal{M}_0 = m')$ . We have:

$$Q(\mathcal{M}_0 = m | \mathcal{M}_\tau = m') = \frac{Q(\mathcal{M}_0 = m, \mathcal{M}_\tau = m')}{Q(\mathcal{M}_\tau = m')} .$$

The denominator is simply the probability of observing the memory state  $m'$  at the end of the forward process:

$$\begin{aligned} Q(\mathcal{M}_\tau = m') &= \int_{\vec{m}} d\vec{s} \Pr_{\vec{x}}(\vec{\mathcal{S}} = \mathbf{\mathfrak{H}} \vec{s}, \mathcal{M}_0 = m'^\dagger | \mathcal{S}_0 = s_\tau^\dagger) \mu_\tau(s_\tau) \\ &= \int_{\vec{m}} d\vec{s} \Pr_{\vec{x}}(\vec{\mathcal{S}} = \mathbf{\mathfrak{H}} \vec{s} | \mathcal{S}_0 = s_\tau^\dagger) \mu_\tau(s_\tau) \\ &= \int_{m'} ds_\tau \mu_\tau(s_\tau) \\ &= P(\mathcal{M}_\tau = m') . \end{aligned}$$

Since the process begins and ends in a metastable distribution, the probability of observing a particular microstate given some memory state is the same as the probability of observing that microstate given that the system has locally equilibrated to that memory state. And, since the process is time symmetric, the starting and ending local equilibrium distributions are the same. That is, letting  $\pi_t^{(m')}(s)$  denote the local equilibrium probability of observing microstate  $s$  given memory state  $m'$  at time  $t$ :

$$P(\mathcal{S}_\tau = s_\tau | \mathcal{M}_\tau = m') = \pi_\tau^{(m')}(s_\tau) = \pi_0^{(m')}(s_\tau) .$$

Additionally:

$$\begin{aligned} P(\mathcal{S}_\tau = s_\tau | \mathcal{M}_\tau = m') &= \frac{P(\mathcal{S}_\tau = s_\tau) [\mathbb{1}_{s_\tau \in m'}]}{P(\mathcal{M}_\tau = m')} \\ &= \frac{\mu_\tau(s_\tau) [\mathbb{1}_{s_\tau \in m'}]}{P(\mathcal{M}_\tau = m')} . \end{aligned}$$

We then have:

$$\begin{aligned} Q(\mathcal{M}_0 = m | \mathcal{M}_\tau = m') &= \frac{\int_{s_0 \in m, s_\tau \in m'} d\vec{s} \Pr_{\vec{x}}(\vec{\mathcal{S}} = \mathbf{\mathfrak{H}} \vec{s}, \mathcal{M}_0 = m'^\dagger, \mathcal{M}_\tau = m^\dagger | \mathcal{S}_0 = s_\tau^\dagger) \mu_\tau(s_\tau)}{P(\mathcal{M}_\tau = m')} \\ &= \int_{s_0 \in m, s_\tau \in m'} ds_0 ds_\tau \Pr_{\vec{x}}(\mathcal{S}_\tau = s_0^\dagger | \mathcal{S}_0 = s_\tau^\dagger) \pi_0^{(m')}(s_\tau) \\ &= \int_{s^\dagger \in m, s'^\dagger \in m'} ds ds' \Pr_{\vec{x}}(\mathcal{S}_\tau = s | \mathcal{S}_0 = s') \pi_0^{(m')}(s'^\dagger) \\ &= \int_{s \in m^\dagger, s' \in m'^\dagger} ds ds' \Pr_{\vec{x}}(\mathcal{S}_\tau = s | \mathcal{S}_0 = s') \pi_0^{(m')}(s'^\dagger) . \end{aligned}$$

Note that in the second to last line we simply changed the symbols denoting the variables of integration.

We now invoke our assumption that the memory partitions are chosen to be time-reversal invariant, such that  $m^\dagger = m$  for all  $m \in \mathcal{M}$ . If we furthermore assume that the equilibrium probability of a microstate does not change under time reversal, then time-reversal symmetry of the memory states implies that the local equilibrium probability of a

microstate does not change either:

$$\pi_0^{(m)}(s) = \pi_0^{(m)}(s^\dagger) .$$

We therefore have:

$$\begin{aligned} Q(\mathcal{M}_0 = m | \mathcal{M}_\tau = m') &= \int_{s \in m, s' \in m'} ds ds' \Pr_{\vec{x}}(\mathcal{S}_\tau = s | \mathcal{S}_0 = s') \pi_0^{(m')}(s') \\ &= \int_{\vec{m}} d\vec{s} \Pr_{\vec{x}}(\vec{\mathcal{S}} = \vec{s}, \mathcal{M}_\tau = m | \mathcal{S}_0 = s_0) \pi_0^{(m')}(s_0) \\ &= P(\mathcal{M}_\tau = m | \mathcal{M}_0 = m') . \end{aligned}$$

This allows us to rewrite the memory-state trajectory probability under the reverse process in terms of forward process probabilities:

$$\begin{aligned} Q(\mathcal{M}_0 = m, \mathcal{M}_\tau = m') &= Q(\mathcal{M}_\tau = m') Q(\mathcal{M}_0 = m | \mathcal{M}_\tau = m') \\ &= P(\mathcal{M}_\tau = m') P(\mathcal{M}_\tau = m | \mathcal{M}_0 = m') \\ &= \mu'_\tau(m') P(\mathcal{M}_\tau = m | \mathcal{M}_0 = m') . \end{aligned}$$

We now show that  $D_{\text{KL}}[P || Q]_{\vec{\mathcal{M}}}$  equals the right hand side of Eq. (2), completing the proof:

$$\begin{aligned} D_{\text{KL}}[P || Q]_{\vec{\mathcal{M}}} &= \sum_{\vec{m} \in \vec{\mathcal{M}}} P(\vec{m}) \ln \frac{P(\vec{m})}{Q(\vec{m})} \\ &= \sum_{m, m' \in \mathcal{M}} P(m, m') \ln \frac{P(m, m')}{Q(m, m')} \\ &= \sum_{m, m' \in \mathcal{M}} P(m, m') \ln \frac{\mu'_0(m) P(m' | m)}{\mu'_\tau(m') P(m | m')} \\ &= \sum_{m, m' \in \mathcal{M}} P(m, m') \ln \mu'_0(m) - \sum_{m, m' \in \mathcal{M}} P(m, m') \ln \mu'_\tau(m') + \sum_{m, m' \in \mathcal{M}} P(m, m') \ln \frac{P(m' | m)}{P(m | m')} \\ &= \sum_{m \in \mathcal{M}} \mu'_0(m) \ln \mu'_0(m) - \sum_{m' \in \mathcal{M}} \mu'_\tau(m') \ln \mu'_\tau(m') + \sum_{m \in \mathcal{M}} \mu'_0(m) \sum_{m' \in \mathcal{M}} P(m' | m) \ln \frac{P(m' | m)}{P(m | m')} \\ &= -H(\mathcal{M}_0) + H(\mathcal{M}_\tau) + \sum_{m \in \mathcal{M}} \mu'_0(m) \sum_{m' \in \mathcal{M}} d(m, m') \\ &= \Delta H(\mathcal{M}_\tau) + \sum_{m \in \mathcal{M}} \mu'_0(m) \sum_{m' \in \mathcal{M}} d(m, m') , \end{aligned}$$

where  $H(\mathcal{M}_t)$  is the entropy of the memory state at time  $t$ .

## II. LANGEVIN SIMULATIONS OF ERASURE

To help simulate a wide variety of protocols, we first nondimensionalize the equations of motion, using variables:

$$\begin{aligned} q' &= \sqrt{\frac{2a}{b_0}} q , \quad t' = \frac{2ak_B T}{b_0 \lambda} t , \quad v' = \frac{\lambda}{k_B T} \sqrt{\frac{b_0}{2a}} v , \quad \text{and} \\ V' &= \frac{1}{k_B T} V . \end{aligned}$$

Note that the position scale  $\sqrt{b_0/2a}$  is the distance from zero to either well minima in the default potential  $V(\cdot, 0) = V(\cdot, \tau)$ . Substitution then provides the following nondimensional equations of motion:

$$\begin{aligned} dq' &= v' dt' \\ m' dv' &= - \left( \frac{\partial}{\partial q'} V'(q', t') + v' \right) dt' + \sqrt{2} r(t') \sqrt{dt'} , \end{aligned}$$

with:

$$m' = \frac{2amk_{\text{B}}T}{b_0\lambda^2},$$

which is the first nondimensional parameter to specify an erasure process.

The nondimensional potential can be expressed as:

$$V'(q', t') = \alpha \left( q'^4 - 2b'_f(t')q'^2 + \zeta c'_f(t')q' \right),$$

where:

$$\alpha = \frac{b_0^2}{4ak_{\text{B}}T} \quad \text{and} \quad \zeta = c_0 \sqrt{\frac{2a}{b_0^3}}$$

are two more nondimensional parameters to specify and:

$$b'_f(t') = b_f \left( \frac{2ak_{\text{B}}T}{b_0\lambda} t \right) \quad \text{and} \quad c'_f(t') = c_f \left( \frac{2ak_{\text{B}}T}{b_0\lambda} t \right)$$

simply express  $b_f$  and  $c_f$  with the nondimensional time as input. The fourth and final nondimensional parameter is the nondimensional total time:

$$\tau' = \frac{2ak_{\text{B}}T}{b_0\lambda} \tau.$$

To explore the space of possible underdamped erasure dynamics, we simulate 735 different protocols, determined by all combinations of the following values for the four nondimensional parameters:  $m' \in \{0.25, 1.0, 4.0\}$ ,  $\alpha \in \{2, 4, 7, 10, 12\}$ ,  $\zeta \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$ ,

and  $\tau' \in \{4, 8, 16, 32, 64, 128, 256\}$ .  $10^5$  trials of each parameter set were simulated.

For the simulations of Figs. 4 and 5, we set  $m' = 1$ ,  $\alpha = 7$ ,  $\zeta = 0.4$ , and  $\tau' = 100$ . Figure 6 shows that the (error, work) pairs obtained for these various dynamics fill in the region allowed by our time-symmetric bounds. These bounds can indeed be tight, but it is quite possible to waste more energy if the computation is not tuned for energetic efficiency.

To update particle position and velocity each time step, we used fourth-order Runge–Kutta integration for the deterministic portion of the equations of motion and a simple Euler method in combination with a Gaussian number generator for the stochastic portion. To determine the time step size, we considered a range of possible time steps for 81 of the possible 735 parameter sets and looked for convergence of the sampled average works and maximum errors  $\epsilon$ , again using  $10^5$  trials per parameter set.

The maximum errors were stable over the whole range of tested step sizes. Exploring decreasing step sizes, a step size of 0.0025 was chosen since there the average works stopped fluctuating within  $5\sigma$  of their statistical errors for all 81 parameter sets. The error bars presented for the average works in Fig. 6 were then generously set to be 5 times the estimated statistical errors, which were each obtained by dividing the sampled standard deviation by the square root of the number of trials. Error bars for the maximum errors were set to be the statistical errors of  $\epsilon_{\text{L}}$  or  $\epsilon_{\text{R}}$ , depending on which was the maximum, whose statistical errors were obtained by assuming binomial statistics.

- 
- [1] R. Landauer. Irreversibility and heat generation in the computing process. *IBM J. Res. Develop.*, 5(3):183–191, 1961. [1](#), [2](#), [5](#), [6](#)
  - [2] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa. Thermodynamics of information. *Nature Physics*, 11(2):131–139, February 2015. [1](#)
  - [3] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Identifying functional thermodynamics in autonomous Maxwellian ratchets. *New J. Physics*, 18:023049, 2016. [1](#), [2](#)
  - [4] A. Gomez-Marin, J. M. R. Parrondo, and C. Van den Broeck. Lower bounds on dissipation upon coarse graining. *Phys. Rev. E*, 78(1):011107, 2008. [1](#)
  - [5] S. Deffner and C. Jarzynski. Information processing and the second law of thermodynamics: An inclusive, Hamiltonian approach. *Phys. Rev. X*, 3:041003, 2013. [1](#), [3](#), [7](#), [8](#)
  - [6] J. Li, J. M. Horowitz, T. R. Gingrich, and N. Fakhri. Quantifying dissipation using fluctuating currents. *Nature Comm.*, 10(1):1–9, 2019. [1](#)
  - [7] S. Still. Thermodynamic cost and benefit of memory. *Phys. Rev. Lett.*, 124(5):050601, 2020. [1](#)
  - [8] M. Gopalkrishnan. A cost/speed/reliability tradeoff to erasing. In C. S. Calude and M. J. Dinneen, editors, *Unconventional Computation and Natural Computation*, volume 9252 of *Lecture Notes in Computer Science*, pages 192–201, Berlin, 2015. Springer-Verlag. [1](#)
  - [9] S. Lahiri, J. Sohl-Dickstein, and S. Ganguli. A universal tradeoff between power, precision and speed in physical communication. *arXiv:1603.07758*. [2](#)
  - [10] A. B. Boyd, A. Patra, C. Jarzynski, and J. P. Crutchfield. Shortcuts to thermodynamic computing: The cost of fast and faithful information processing. *arXiv:1812.11241*. [1](#), [2](#)
  - [11] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Thermodynamics of modularity: Structural costs beyond the landauer bound. *Phys. Rev. X*, 8(031036), 2018. [1](#), [2](#)

- [12] P. M. Riechers, A. B. Boyd, G. W. Wimsatt, and J. P. Crutchfield. Balancing error and dissipation in computing. *Phys. Rev. Res.*, 2(3):033524, 2020. 2, 3, 4, 10, 11
- [13] A. Kolchinsky and D. H. Wolpert. Dependence of dissipation on the initial distribution over states. *J. Stat. Mech. Th. Exp.*, 2017(8):083202, 2017. 2
- [14] P. M. Riechers. Transforming metastable memories: The nonequilibrium thermodynamics of computation. In D. Wolpert, C. Kempes, P. Stadler, and J. Grochow, editors, *The Energetics of Computing in Life and Machines*. SFI Press, 2019. 2
- [15] A. B. Boyd, D. Mandal, P. M. Riechers, and J. P. Crutchfield. Transient dissipation and structural costs of physical information transduction. *Phys. Rev. Lett.*, 118:220602, 2017. 2
- [16] P. M. Riechers and J. P. Crutchfield. Fluctuations when driving between nonequilibrium steady states. *J. Stat. Phys.*, 168(4):873–918, 2017. 2
- [17] S. Loomis and J. P. Crutchfield. Thermodynamically-efficient local computation and the inefficiency of quantum memory compression. *Phys. Rev. Res.*, 2(2):023039, 2019. 2
- [18] Technology Working Group. The International Technology Roadmap for Semiconductors 2.0: 2015, Executive Summary. Technical report, Semiconductor Industry Association, 2015. 2
- [19] T. Conte et al. Thermodynamic computing. *arxiv:1911.01968*. 2
- [20] Technology Working Group. The International Roadmap for Devices and Systems: 2020, Executive Summary. Technical report, Institute of Electrical and Electronics Engineers, 2020. 2
- [21] Technology Working Group. The International Roadmap for Devices and Systems: 2020, More Moore. Technical report, Institute of Electrical and Electronics Engineers, 2020.
- [22] Technology Working Group. The International Roadmap for Devices and Systems: 2020, Beyond CMOS. Technical report, Institute of Electrical and Electronics Engineers, 2020.
- [23] J. Shalf. The future of computing beyond Moore’s law. *Phil. Trans. Roy. Soc.*, 378:20190061, 2020. 2
- [24] R. Klages, W. Just, and C. Jarzynski, editors. *Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond*. Wiley, New York, 2013. 1
- [25] T. Sagawa. Thermodynamics of information processing in small systems. *Prog. Theo. Phys.*, 127(1):1–56, 2012. 2
- [26] A. Berut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz. Experimental verification of Landauer’s principle linking information and thermodynamics. *Nature*, 483:187–190, 2012.
- [27] J. L. England. Dissipative adaptation in driven self-assembly. *Nature Nanotech.*, 10(11):919, 2015.
- [28] O.-P. Saira, M. H. Matheny, R. Katti, W. Fon, G. Wimsatt, S. Han, J. P. Crutchfield, and M. L. Roukes. Nonequilibrium thermodynamics of erasure with superconducting flux logic. *Phys. Rev. Res.*, 2:013249, 2020. 2
- [29] C. H. Bennett. Thermodynamics of computation—A review. *Intl. J. Theo. Phys.*, 21:905, 1982. 2
- [30] T. Sagawa. Thermodynamic and logical reversibilities revisited. *J. Stat. Mech. Th. Exp.*, 2014(3):P03025, 2014. 2
- [31] P. R. Zulkowski and M. R. DeWeese. Optimal finite-time erasure of a classical bit. *Phys. Rev. E*, 89:052140, 2014. 2
- [32] P. R. Zulkowski and M. R. DeWeese. Optimal protocols for driven quantum systems. 2014. arXiv:1506.03864.
- [33] P. R. Zulkowski and M. R. DeWeese. Optimal control of overdamped systems. *Phys. Rev. E*, 92(3):032117, 2015. 2
- [34] E. Aurell, K. Gawędzki, C. Mejía-Monasterio, R. Mohayaei, and P. Muratore-Ginanneschi. Refined second law of thermodynamics for fast random processes. *J. Stat. Physics*, 147(3):487–505, 2012. 2
- [35] K. Proesmans, J. Ehrich, and J. Bechhoefer. Finite-time Landauer principle. *Phys. Rev. Lett.*, 125(10):100602, 2020.
- [36] K. Proesmans, J. Ehrich, and J. Bechhoefer. Optimal finite-time bit erasure under full control. *Phys. Rev. E*, 102(3):032105, 2020. 2
- [37] C. Jarzynski. Hamiltonian derivation of a detailed fluctuation theorem. *J. Stat. Phys.*, 98(1-2):77–102, 2000. 3
- [38] G. E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, 60:2721, 1999. 3
- [39] G. Lan, P. Sartori, S. Neumann, V. Sourjik, and Y. Tu. The energy–speed–accuracy trade-off in sensory adaptation. *Nature Physics*, 8(5):422, 2012. 5
- [40] P. M. Riechers and J. P. Crutchfield. Fluctuations when driving between nonequilibrium steady states. *J. Stat. Physics*, 168(4):873–918, 2017. 5
- [41] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, second edition, 2006. 5
- [42] B. Schroeder, E. Pinheiro, and W.-D. Weber. DRAM errors in the wild: A large-scale field study. *SIGMETRICS/Performance’09, Seattle, WA*, pages 1–12, 2009. 6
- [43] Y. Jun, M. Gavrilov, and J. Bechhoefer. High-precision test of Landauer’s principle. *Phys. Rev. Lett.*, 113:190601, 2014. 6
- [44] 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. 6
- [45] R. Dillenschneider and E. Lutz. Memory erasure in small systems. *Phys. Rev. Lett.*, 102:210601, May 2009. 10
- [46] Y. Jun, M. Gavrilov, and J. Bechhoefer. High-precision test of Landauer’s principle in a feedback trap. *Phys. Rev. Lett.*, 113:190601, Nov 2014. 10
- [47] D. Reeb and M. M. Wolf. An improved Landauer principle with finite-size corrections. *New J. Physics*, 16(10):103011, 2014. 10
- [48] A. M. Timpanaro, J. P. Santos, and G. T. Landi. Landauer’s principle at zero temperature. *Phys. Rev. Lett.*,

- 124(24):240601, 2020. [10](#)
- [49] 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(16):160602, 2020. [10](#)
- [50] S. Sheng, T. Herpich, G. Diana, and M. Esposito. Thermodynamics of majority-logic decoding in information erasure. *Entropy*, 21(3):284, 2019.
- [51] K. Proesmans and J. Bechhoefer. Erasing a majority-logic bit. *arXiv:2010.15885*, 2020. [10](#)
- [52] R. Landauer. Private communication with J. P. Crutchfield, 1981. [10](#)