**Title**

Balancing error and dissipation in computing

**Permalink**

**Journal**

**ISSN**

**Authors**

Riechers, PM
Boyd, AB
Wimsatt, GW
et al.

**Publication Date**

**DOI**

Peer reviewed

# Balancing Error and Dissipation in Computing

Paul M. Riechers[*] and Alexander B. Boyd[†]

*Complexity Institute and School of Physical and Mathematical Sciences,*
*Nanyang Technological University, 637371 Singapore, Singapore*

Gregory W. Wimsatt[‡] and James P. Crutchfield[§]

*Complexity Sciences Center and Physics Department,*
*University of California at Davis, One Shields Avenue, Davis, CA 95616*

Modern digital electronics support remarkably reliable computing, especially given the challenge of controlling nanoscale logical components that interact in fluctuating environments. However, we demonstrate that the high-reliability limit is subject to a fundamental error–energy-efficiency tradeoff that arises from time-symmetric control: Requiring a low probability of error causes energy consumption to diverge as logarithm of the inverse error rate for nonreciprocal logical transitions. The *reciprocity* (*self*-invertibility) of a computation is a stricter condition for thermodynamic efficiency than logical reversibility (invertibility), the latter being the root of Landauer's work bound on erasing information. Beyond engineered computation, the results identify a generic error–dissipation tradeoff in steady-state transformations of genetic information carried out by biological organisms. The lesson is that computation under time-symmetric control cannot reach, and is often far above, the Landauer limit. In this way, time-asymmetry becomes a design principle for thermodynamically efficient computing.

## CONTENTS

## I. INTRODUCTION

Tantalizingly, the thermodynamics of computation tells us that information processing can be achieved with zero energy dissipation ... if one has sufficient control over a system's microscopic degrees of freedom and can endure the quasistatic limit of infinitely-slow processing [1–5]. To be useful, though, computation must be performed in finite time. Unfortunately, this requires additional work and guarantees the investment is lost via dissipation. This state of affairs poses a grand challenge to thermodynamic computing: Identify control protocols that reliably drive a system between memory states according to a desired computation in finite time and with minimal dissipation. Failing an answer, the fundamental physical limits on computation remain elusive.

In point of fact, contemporary finite-time thermodynamics predicts that energy-efficient protocols of duration $\tau$ entail dissipation that scales as $\tau^{-1}$ [6–11]. That is, reliable computation could be performed with arbitrarily little dissipation at the cost of arbitrarily slow processing.

[*] pmriechers@gmail.com
[†] abboyd@ucdavis.edu
[‡] gwwimsatt@ucdavis.edu
[§] chaos@ucdavis.edu

An obvious and common implementation of finite-time computation is to ramp up a set of forces[1] that transform previously-metastable memories according to the computation's input–output map. At an appropriate later time, the forces are ramped down (mirroring the ramp-up procedure) so that the computing system returns to its resting state, while new memories are again stored robustly in metastable configurations. Such implementations represent control protocols that are symmetric in time.

Such time-symmetric protocols are common. Most notably, the primary signal controlling information processing in contemporary microprocessors is a time-symmetric clock-voltage signal—a several gigahertz square wave that orchestrates all transformations of the computer's logic and memory components [12, 13].

Time-symmetric protocols are common in other settings that seek reliable transformations. For example, recent explorations of synthetic molecular machines externally drive molecular rotors with sinusoidal (time-symmetric) electric fields [14]. And, autonomous synthetic molecular machines are known to reliably produce steady-state rotations embedded in a time-*invariant* nonequilibrium environment [15]—a special case of time-symmetric control. They achieve this through chemical catalysis, much as natural molecular machines operate in vivo. In vivo, reliable genetic transformations of DNA (and reliable directional motion of molecular motors) occur in environments with very low Reynolds number, where there are no inertial variables to freely exert a time-asymmetric influence [16, 17]. More broadly, breaking time symmetry in Brownian motion *requires* significant dissipation [18]. And so, in both the biological and engineered worlds, simple time-symmetric protocol design—turning on and then turning off an interaction—seems to offer an energetically-efficient and reliable way to implement change.

The goal, naturally enough, is to implement information processing in ways that require no more than the minimal work exertion set by Landauer's logical-irreversibility bound [19]: $W \geq k_{\mathrm{B}}T \ln 2$. One might even hope that this work could then be recycled rather than dissipated. Indeed, as appreciated recently, changes in the nonequilibrium addition to free energy can be leveraged to implement logically-irreversible computations in a thermodynamically reversible manner [2–5]. The vision is of a future of hyper-efficient computing using orders-of-magnitude less energy than currently.

However, there is a wrinkle in this optimism. The following shows that *all* time-symmetric protocols for transforming metastable memories engender an irreducible trade-off

between computational accuracy and energy efficiency. Specifically, in the limit of highly-reliable computing with vanishingly-small error probability $\epsilon$, the minimal dissipation under time-symmetric controls diverges as $-\ln \epsilon$ with a coefficient proportional to the computation's *nonreciprocity*—the nonself-invertibility of its memory-state transitions.

That is, for reliable time-symmetric implementations of a deterministic computation $\mathcal{C}$, which maps each memory state $m \mapsto \mathcal{C}(m)$, we show that the minimal work above the change in local-equilibrium free energy is $k_{\mathrm{B}}T \ln(1/\epsilon)$ whenever $\mathcal{C}(\mathcal{C}(m)) \neq m$. That is, work $k_{\mathrm{B}}T \ln(1/\epsilon)$ is required whenever the computation iterated twice does not return to the original memory state. Moreover, there is a correction when time-odd variables store memory that highlights the unique advantages of both magnetic and conformational memories.

In this way, nonreciprocity identifies a dominant cost of thermodynamic computing for the broad class of time-symmetric control protocols. While future work addresses more general thermodynamic implications of reliable computation, since time-symmetric control protocols are common and often unavoidable in key applications, they are the subject of our error-dissipation tradeoff analysis here.

Our first step revisits the basics of dissipation in thermodynamic systems and the cost of partial macroscopic knowledge, when the entire microstate cannot be observed or controlled. The dissipation costs are then recast to apply to thermodynamic computing generally and, in particular, to information storage and processing in metastable mesoscopic states. We then analyze the role of time symmetries in control protocols and give a thermodynamic accounting. Dissipation scaling with error level is derived and applied in the limit of highly-reliable computing. This is then used to analyze dissipation in erasure, logic gates, and biological systems. We conclude, briefly comparing recent work on control restrictions and noting directions for further exploration. Appendices provide details underlying the theory.

## II. COMPUTATIONAL DISSIPATION

To start, consider any physical realization of a memory device with a system Hamiltonian $\mathcal{H}_x$ parametrized by control $x \in \mathcal{X}$. For example, $x$ could correspond to an applied electromagnetic field, a vector of quantities specifying a potential energy landscape, or a set of piston positions. The Hamiltonian $\mathcal{H}_x$ determines the system's instantaneous energies $\{E_x(s)\}_{s \in \mathcal{S}}$, where $\mathcal{S}$ is the set of system microstates. We assume that the system is in contact with an effectively memoryless heat bath at temperature $T$ that enables the system to relax to both local and global equilibria. The instantaneous stochastic microstate dynamics are completely determined by the

---

[1] For example, a set of gate voltages may be applied or, by any other method, a potential energy landscape may be altered to temporarily couple and bias a set of initially-independent metastable potential minima.

system's instantaneous Hamiltonian and the system's interaction with the heat bath.

Work is performed by driving the system via a *control protocol* $x_{0:\tau} = x_0 \cdots x_\tau$, which is the trajectory of the control parameters from time 0 to $\tau$. The work $W$ measures the accumulated change in the energy of occupied microstates—energy that was supplied by the controller. In the following, we are especially interested in *computations* implemented by the control protocol and in the associated work cost.

## A. Background

Understanding the relationship between work costs and computations requires tracking the energetics of microstate trajectories $s_{0:\tau} = s_0 \cdots s_\tau$ of the system from time 0 to $\tau$. Via the implied dynamics set by the time-dependent Hamiltonian $\mathcal{H}_{x_t}$ and the coupling with the bath at temperature $T$, the initial state $s_0$ and the control protocol $x_{0:\tau}$ determine the probability of microstate trajectories:

$$\Pr_{x_{0:\tau}}(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{S}_0 = s_0)$$
$$\equiv \Pr(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{S}_0 = s_0, X_{0:\tau} = x_{0:\tau}) , \quad (1)$$

where $\mathcal{S}_t$ is the random variable for the microstate at time $t$ and $\mathcal{S}_{0:\tau}$ is the random variable chain for the full state trajectory. The subscripted probability $\Pr_{x_{0:\tau}}(\cdot)$ denotes the probability distribution *induced* by the driving protocol $x_{0:\tau}$. This is the same as the probability *conditioned* on the driving protocol, as described by Eq. (1).

Hamiltonian control implies microscopic reversibility of the instantaneous system–bath dynamics. This means state trajectories that release energy into the heat bath ($Q > 0$) are exponentially more likely than the time-reversed microstate trajectory of the system (that absorb heat), if the protocol were run in reverse [20, 21]. In the Markovian limit:[2]

$$\frac{\Pr_{x_{0:\tau}}(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{S}_0 = s_0)}{\Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})}(\mathcal{S}_{0:\tau} = \boldsymbol{\mathcal{A}}(s_{0:\tau}) | \mathcal{S}_0 = s_\tau^\dagger)} = e^{\beta Q(s_{0:\tau}, x_{0:\tau})} , \quad (2)$$

where $\beta = (k_{\mathrm{B}} T)^{-1}$ and $k_{\mathrm{B}}$ is Boltzmann's constant. $\boldsymbol{\mathcal{A}}$ denotes *time-reversal*, including not only reversal of the time ordering, but also conjugation of time-odd variables such as momentum and spin: $s_0 \ldots s_\tau \mapsto s_\tau^\dagger \ldots s_0^\dagger$. For example, if the microstate $s = (\vec{q}, \vec{\wp})$ is a collection of spatial $\vec{q}$ and momentum $\vec{\wp}$ degrees of freedom, then the

conjugation simply flips all of the momentum degrees of freedom $s^\dagger = (\vec{q}, -\vec{\wp})$.

Thermodynamic irreversibility is quantified by *entropy production* $\Sigma$ which, by the second law, is nonnegative on average: $\langle \Sigma \rangle \geq 0$. In our setting, $\Sigma$ is simply the net change in the component entropies of both the system and the bath.

More specifically, entropy production decomposes into the change in heat-bath entropy $Q(s_{0:\tau}, x_{0:\tau})/T$ beyond any compensating reduction in the system's microstate entropy. The system's instantaneous microstate entropy is given by the nonequilibrium *surprisal* $-k_{\mathrm{B}} \ln \boldsymbol{\mu}_t(s_t)$, where $\boldsymbol{\mu}_t(s_t) = \Pr_{x_{-\infty:t}}(\mathcal{S}_t = s_t)$ is the current microstate's probability given the *entire history* of preparation and driving [20, 22, 23].[3] Accordingly, the entropy production is:

$$\Sigma = \frac{1}{T} Q(s_{0:\tau}, x_{0:\tau}) + \Delta\big(-k_{\mathrm{B}} \ln \boldsymbol{\mu}_t(s_t)\big) ,$$

where $\Delta$ indicates the change from time 0 to $\tau$. Specifically, $\boldsymbol{\mu}_0$ is the distribution over microstates given the system's initial preparation and $\boldsymbol{\mu}_\tau$ is the time-evolved version of $\boldsymbol{\mu}_0$ under the influence of the driving $x_{0:\tau}$.

This results in the trial-specific entropy production [20, 21]:

$$\Sigma = \frac{1}{T} Q(s_{0:\tau}, x_{0:\tau}) + k_{\mathrm{B}} \ln \frac{\boldsymbol{\mu}_0(s_0)}{\boldsymbol{\mu}_\tau(s_\tau)}$$
$$= k_{\mathrm{B}} \ln \left( \frac{\Pr_{x_{0:\tau}}(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{S}_0 = s_0) \boldsymbol{\mu}_0(s_0)}{\Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})}(\mathcal{S}_{0:\tau} = \boldsymbol{\mathcal{A}}(s_{0:\tau}) | \mathcal{S}_0 = s_\tau^\dagger) \boldsymbol{\mu}_\tau(s_\tau)} \right)$$
$$= k_{\mathrm{B}} \ln \left( \frac{\Pr_{x_{0:\tau}}(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{S}_0 \sim \boldsymbol{\mu}_0)}{\Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})}(\mathcal{S}_{0:\tau} = \boldsymbol{\mathcal{A}}(s_{0:\tau}) | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger)} \right) , \quad (3)$$

where $\boldsymbol{\mu}_\tau^\dagger$ is the distribution such that $\boldsymbol{\mu}_\tau^\dagger(s) = \boldsymbol{\mu}_\tau(s^\dagger)$. The condition $\mathcal{S}_0 \sim \boldsymbol{\mu}$ means the random variable $\mathcal{S}_0$ is distributed according to $\boldsymbol{\mu}$, as in Ref. [24]. We used the fact that $\boldsymbol{\mu}_\tau(s) = \boldsymbol{\mu}_\tau^\dagger(s^\dagger)$ in going from the second to the third line.

When all influences driving the system away from equilibrium are controlled changes to the system Hamiltonian, the total entropy production is proportional to the *dissipated work* $W_{\mathrm{diss}}$. This is the accumulation of lost opportunities to extract work:

$$W_{\mathrm{diss}} = T\Sigma .$$

---

[2] In the more general non-Markovian case, Eq. (2) becomes an *in*equality which nevertheless still upholds our main results bounding computational dissipation.

[3] Hence, $\boldsymbol{\mu}_t(s_t)$ is the expected probability of being in the actual microstate $s_t$ over many trials where the system is prepared and driven in exactly the same way. The fact that this distribution is not $\delta$-distributed is due to the stochasticity induced by the interaction with the heat bath (and any other uncontrollable degrees of freedom).

The entropy production and dissipated work both measure a given computation's efficiency, as they both quantify the degree to which more work was done than necessary by an unrestricted controller.

The *expected* dissipated work then quantifies the difference between the average amount of work done beyond the change in nonequilibrium free energy [25]:

$$\langle W_{\mathrm{diss}} \rangle = \langle W \rangle - \Delta \mathcal{F} \ . \tag{4}$$

This is the average work that has been irreversibly lost, since the nonequilibrium free energy $\mathcal{F}$ is the expected amount of energy that possibly could be extracted as work [4].

Calculating the dissipated work from these two quantities seems to require explicit knowledge of the Hamiltonian, since the work is the integrated change in the energy due to changes in control:

$$\langle W \rangle = \sum_s \int_0^\tau dt \, \boldsymbol{\mu}_t(s) \, \frac{\partial E_{x_t}(s)}{\partial x_t} \frac{dx_t}{dt} \ ,$$

and the nonequilibrium free energy is the difference between the average energy and the microstate entropy [4]:

$$\mathcal{F}(t) = \langle \mathcal{H}_{x_t} \rangle - k_{\mathrm{B}} T \, \mathrm{H}(\mathcal{S}_t) \ ,$$

where $\mathrm{H}(Z) = - \sum_z \Pr(Z = z) \ln \Pr(Z = z)$ is the Shannon entropy of the driven system in nats.[4]

However, due to Eq. (3), we can calculate the dissipated work for a control protocol using only the probability of forward trajectories under forward driving $\rho(\mathcal{S}_{0:\tau} = s_{0:\tau}) = \Pr_{x_{0:\tau}} \big( \mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{S}_0 \sim \boldsymbol{\mu}_0 \big)$ and reverse microstate trajectories under reverse driving $\rho^R(\mathcal{S}_{0:\tau} = s_{0:\tau}) = \Pr_{\boldsymbol{\mathit{A}}(x_{0:\tau})} \big( \mathcal{S}_{0:\tau} = \boldsymbol{\mathit{A}}(s_{0:\tau}) | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger \big)$. Averaging over forward trajectories produces a relative entropy:

$$\beta \langle W_{\mathrm{diss}} \rangle = \sum_{s_{0:\tau}} \rho(\mathcal{S}_{0:\tau} = s_{0:\tau}) \ln \frac{\rho(\mathcal{S}_{0:\tau} = s_{0:\tau})}{\rho^R(\mathcal{S}_{0:\tau} = s_{0:\tau})}$$
$$= \mathrm{D}_{\mathrm{KL}} \left[ \rho(\mathcal{S}_{0:\tau}) \, || \, \rho^R(\mathcal{S}_{0:\tau}) \right] \ , \tag{5}$$

where $\mathrm{D}_{\mathrm{KL}}[\cdot \, || \, \cdot]$ is the Kullback–Leibler divergence.

Eq. (5) appears superficially similar to many previous related results, including Refs. [26–30]. However these previous results all required that the system starts in a

steady-state distribution (either equilibrium or a nonequilibrium steady state), which precludes the storage of general initial memories. Eq. (5) overcomes that previous limitation via its dependence on $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_\tau^\dagger$—a crucial generalization that will allow us to address the thermodynamics of general memory transformations.

By expressing the dissipated work and entropy production in terms of the relative entropy between forward and reverse trajectories, Eq. (5) provides a method to reconstruct a system's thermodynamics without explicit knowledge of the underlying Hamiltonian mechanics. This only requires knowledge of trajectory probabilities under forward and reverse driving and these can be determined experimentally. However, practical considerations can stand in the way: Often direct observations of a thermodynamic system's underlying microstate trajectories are not available.

Fortunately, via the information processing inequality [24], this powerful relation leads to bounds on dissipated work based on *observed trajectories*, which contain only partial details of the full microstate trajectory. In particular, any observable coarse-graining of the microstate trajectory yields a lower bound on the dissipation [27, 28, 31]. This will allow us to infer new lower bounds on computational dissipation from only partial knowledge related to the logical dynamics of the computation itself.

## B. Thermodynamic Computing

To be usefully manipulated, memories must be physically addressable. Accordingly, we define the set $\boldsymbol{\mathcal{M}}$ of *memory states* as a partition of the set $\boldsymbol{\mathcal{S}}$ of microstates, such that each memory state $m \in \boldsymbol{\mathcal{M}}$ is a subset of the microstates: $m \subset \boldsymbol{\mathcal{S}}$ and $\bigcup m = \boldsymbol{\mathcal{S}}$. In this way, the memory state random variable $\mathcal{M}_t$ at time $t$ is determined by the microstate $\mathcal{S}_t$. Moreover, for memory states to robustly hold memories between computations, we assume that partitioning is such that each memory state corresponds to a state-space region that is effectively autonomous when no computation is being performed, resulting in metastability of the memory.[5] Figure 1 shows a spatial system that has been partitioned into Left (L) and Right (R) memory states, each of which corresponds to a minimum in the energy landscape.

Partitioning microstates into metastable memory states $\boldsymbol{\mathcal{M}}$ introduces an observational channel for monitoring computation. The task of thermodynamic computation is to map an initial memory state $m$ to a final memory state $m'$ according to a given conditional probability

---

[4] In point of fact, knowledge of the Hamiltonian is not strictly necessary if one instead calculates $\langle W_{\mathrm{diss}} \rangle = \langle W_{\mathrm{ex}} \rangle - \Delta \mathcal{F}_{\mathrm{add}}$, where $W_{\mathrm{ex}} = W - \Delta F_x^{\mathrm{eq}}$ and $\mathcal{F}_{\mathrm{add}} = k_{\mathrm{B}} T \, \mathrm{D}_{\mathrm{KL}} [\boldsymbol{\mu}_t \, || \, \boldsymbol{\pi}_{x_t}]$. With knowledge (or observation) of the equilibrium distribution $\boldsymbol{\pi}_x$ associated with each control setting $x$, one can leverage the Boltzmann relationship $\boldsymbol{\pi}_x(s) = e^{-\beta[E_x(s) - F_x^{\mathrm{eq}}]}$ to calculate: $\beta \langle W_{\mathrm{ex}} \rangle = \sum_s \int_0^\tau dt \, \boldsymbol{\mu}_t(s) \frac{\partial}{\partial x_t} \left( - \ln \boldsymbol{\pi}_{x_t}(s) \right) \frac{dx_t}{dt}$. Still, Eq. (5) presents an even more tempting opportunity for calculating $\langle W_{\mathrm{diss}} \rangle$ directly from observed trajectory probabilities.

[5] *During* a computation, though, driving couples memory states to instantiate nontrivial computing.
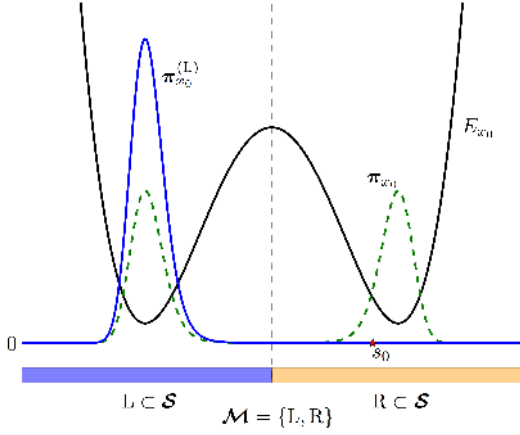
FIG. 1. Bistable memory element: A cross-section (at constant momentum $\wp_0$) of both global and local equilibrium. Initial system state $s_0 = (q_0, \wp_0)$ (red star) lies in a continuum. The energy landscape (black curve) has two symmetric minima that effectively partition the space into robust Left (blue) and Right (yellow) spatial memory states. The stable equilibrium distribution (green dashed curve) is uniformly distributed over the Left and Right states. The restriction of the equilibrium distribution to the Left well (blue curve) is metastable, only locally in equilibrium.
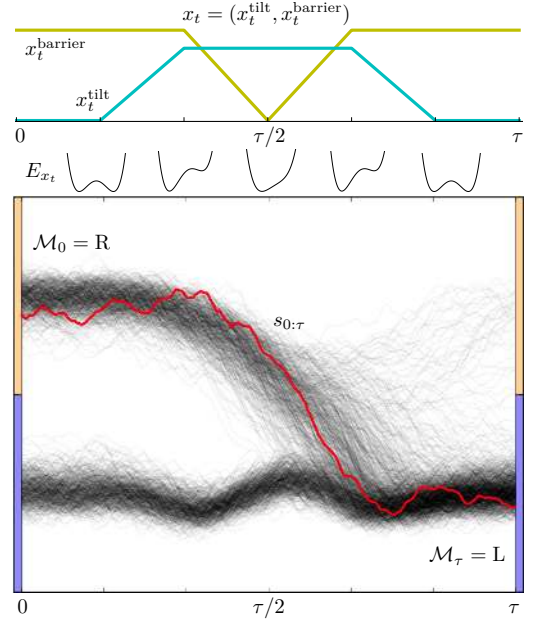


FIG. 2. (Top) Time-symmetric control implementing reliable information processing: erasure to L. (Middle) Controlled potential $E_{x_t}$ at several times during the erasure protocol. (Bottom) Representative behavior via 1000 microstate trajectories in gray which sketches a profile of $\boldsymbol{\mu}_t$. A particular microstate trajectory $s_{0:\tau}$ is highlighted in red. (Note that $q_{0:\tau}$, shown explicitly, implies $\wp_{0:\tau}$ by differentiation and so also implies $s_{0:\tau}$.)

distribution:

$$p(m \to m') = \Pr(\mathcal{M}_\tau = m' | \mathcal{M}_0 = m) .$$

For instance, as shown in Fig. 2 for Fig. 1's system, erasure is composed of two desired transitions with high probabilities—$p(\mathrm{L} \to \mathrm{L}) \approx 1$ and $p(\mathrm{R} \to \mathrm{L}) \approx 1$—between the two metastable memory states. (The latter are labeled L and R according to the common use of left and right minima in a double-well potential landscape to store a bit [32].) A computation occurs due to a particular stochastic trajectory, such as the red path shown in Fig. 2. Naturally, the statistics of the memory state transition probability $p(\mathrm{L} \to \mathrm{L})$, say, come from the ensemble of trajectories. While stochasticity is the rule in thermodynamic computations, the following focuses on nearly-deterministic computations, such as highly-reliable erasure.

To lower-bound the thermodynamic dissipation associated with computation, we consider the coarse graining $s_{0:\tau} \mapsto (m, m')$, where $s_0 \in m$ and $s_\tau \in m'$. This maps the detailed microstate trajectory to the logical transformation it induces between initial and final memory states. For a particular realization $\mathcal{M}_0 = m$ and $\mathcal{M}_\tau = m'$, the time-reversal of the initial and final memory states is: $\boldsymbol{\mathit{A}}(m, m') = (m'^\dagger, m^\dagger)$, where the conjugate $m^\dagger \equiv \{s^\dagger : s \in m\}$. This leads to observable memory transitions:

$$\rho\big((\mathcal{M}_0, \mathcal{M}_\tau) = (m, m')\big)$$
$$= \sum_{s_{0:\tau}} \delta_{s_0 \in m} \delta_{s_\tau \in m'} \, \rho(\mathcal{S}_{0:\tau} = s_{0:\tau})$$
$$= \Pr_{x_{0:\tau}}\big(\mathcal{M}_0 = m, \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0\big) \quad (6)$$

and the reversal probabilities:

$$\rho^R\big((\mathcal{M}_0, \mathcal{M}_\tau) = (m, m')\big)$$
$$= \sum_{s_{0:\tau}} \delta_{s_0 \in m} \delta_{s_\tau \in m'} \, \rho^R(\mathcal{S}_{0:\tau} = s_{0:\tau})$$
$$= \Pr_{\boldsymbol{\mathit{A}}(x_{0:\tau})}\big(\mathcal{M}_0 = m'^\dagger, \mathcal{M}_\tau = m^\dagger | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger\big) . \quad (7)$$

Appendix A uses this trajectory coarse-graining to show that the dissipated work is lower-bounded by a function of the net transition probabilities between memory states:

$$\beta \langle W_{\mathrm{diss}} \rangle = \mathrm{D}_{\mathrm{KL}}\big[\rho(\mathcal{S}_{0:\tau}) \,\big|\big|\, \rho^R(\mathcal{S}_{0:\tau})\big] \quad (8)$$
$$\geq \mathrm{D}_{\mathrm{KL}}\big[\rho(\mathcal{M}_0, \mathcal{M}_\tau) \,\big|\big|\, \rho^R(\mathcal{M}_0, \mathcal{M}_\tau)\big] \quad (9)$$
$$= \Delta \mathrm{H}(\mathcal{M}_t) + \sum_{m,m' \in \boldsymbol{\mathcal{M}}} \boldsymbol{\mu}_0(m) \, d(m, m') , \quad (10)$$

where $\boldsymbol{\mu}_t$ is the same preparation-and-driving-induced probability measure as previously introduced, such that

$\boldsymbol{\mu}_t(m) = \Pr(\mathcal{M}_t = m | \mathcal{S}_t \sim \boldsymbol{\mu}_t) = \sum_{s \in m} \boldsymbol{\mu}_t(s)$, and:

$$\Delta \, \mathrm{H}(\mathcal{M}_t) = \sum_{m \in \boldsymbol{\mathcal{M}}} \Big( \boldsymbol{\mu}_0(m) \ln \boldsymbol{\mu}_0(m) - \boldsymbol{\mu}_\tau(m) \ln \boldsymbol{\mu}_\tau(m) \Big)$$

is the change in Shannon entropy (in nats) of the coarse-grained memory states.

Note that Eq. (10) defined:

$$d(m,m') \equiv \Pr_{x_{0:\tau}} \big( \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0^{(m)} \big)$$

$$\times \ln \left( \frac{\Pr_{x_{0:\tau}} \big( \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0^{(m)} \big)}{\Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \big( \mathcal{M}_\tau = m^\dagger | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^{\dagger(m'^\dagger)} \big)} \right), \quad (11)$$

where $\boldsymbol{\mu}_t^{(m)} \equiv \delta_{\mathcal{S}_t \in m} \boldsymbol{\mu}_t / \boldsymbol{\mu}_t(m)$ is the renormalized microstate distribution $\boldsymbol{\mu}_t$ restricted to memory $m$'s microstates. This compares (i) the probability of transitioning from memory state $m$ to $m'$ to (ii) the probability of returning to $m^\dagger$ upon subsequent momentum-conjugation of the microstate distribution in $m'$ and reversal of the control protocol. One concludes that dissipation is due to statistical irreversibility. At first sight, Eqs. (10) and (11) do not appear to significantly simplify the problem of inferring dissipation from partial knowledge. However, the practical constraints of reliable computation greatly simplify $d(m, m')$, as we will see shortly.

To emphasize, written as either Eq. (9) or (10), the bound applies when the conjugate $m^\dagger \equiv \{ s^\dagger : s \in m \}$ of a memory state $m$ is an element of the memory partition, $m^\dagger \in \boldsymbol{\mathcal{M}}$, for all $m$. (However, see App. A for the more general case.) This occurs, for example, when spatial cells store a memory that is indifferent to rapidly fluctuating momentum degrees of freedom, so that $m = m^\dagger$. A similar simplification occurs for magnetic memory where $s$ and $s^\dagger$ are always in different memory states, such that $m^\dagger \in \boldsymbol{\mathcal{M}}$ although $m \neq m^\dagger$. These two types of memory, illustrated in Fig. 3, lead to notable physical consequences; ones that we explore in a sequel.

Equation (10) is close to yielding a bound on entropy production divorced from knowledge of microstate dynamics. $\Pr_{x_{0:\tau}}(\mathcal{M}_\tau | \mathcal{S}_0 \sim \boldsymbol{\mu}_0^{(m)})$, for instance, can be recovered from observations, given a system's initialization $\boldsymbol{\mu}_0^{(m)}$ and the driving $x_{0:\tau}$. However, for $\Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \big( \mathcal{M}_\tau | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^{\dagger(m)} \big)$, it is unclear how to experimentally prepare the distribution $\boldsymbol{\mu}_\tau^{\dagger(m)}$ in the presence of time-odd variables, like momentum. Fortunately, as we now show, this obstacle can be removed for thermodynamic computations using metastable memories. Ignoring the underlying microstate dynamics and, instead, only tracking the logical dynamics, we bound the work production while computing.
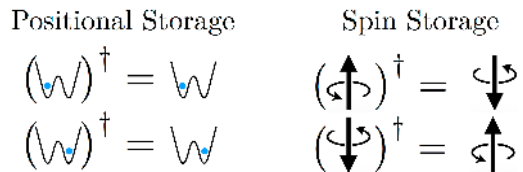


FIG. 3. Common memory elements are either time-reversal invariant (where $m^\dagger = m$) or flip under time-reversal (such that $m^\dagger \neq m$ although $m^\dagger \in \boldsymbol{\mathcal{M}}$).

## C. Metastable Processing

Further simplifying Eq. (10)'s bound on dissipated work turns on recognizing a common property of information processing systems: They hold memory states as metastable nonequilibrium distributions. This has profound thermodynamic implications.

The implied timescale separation requires that relaxing to the local-equilibrium distribution $\boldsymbol{\pi}_x^{(m)}$ within each memory region is fast, while relaxing to the global equilibrium distribution $\boldsymbol{\pi}_x$ is much slower than the timescale of (and between) a system's computational steps. On the one hand, *metastable distributions* are those that correspond (up to some approximation) to a normalized weighted sum of local-equilibrium distributions. On the other, the *global equilibrium distribution* corresponding to any $x$ is the canonical one: $\boldsymbol{\pi}_x(s) = e^{-\beta E_x(s)} / \sum_{s' \in \boldsymbol{\mathcal{S}}} e^{-\beta E_x(s')}$. The system approaches it in the limit of infinite time, if the control setting $x$ is held fixed. In this, the *local-equilibrium distributions* $\boldsymbol{\pi}_x^{(m)}(s) = \delta_{s \in m} \boldsymbol{\pi}_x(s) / \sum_{s' \in m} \boldsymbol{\pi}_x(s')$ are the canonical distributions, associated with each nearly-autonomous region of state space, that local densities approach much more quickly due to timescale separation [25].

If metastable memories are stored robustly between computations, then the memory-state distribution $\boldsymbol{\mu}_t(m)$ at times $t = 0$ and $t = \tau$ contain almost all the information about the microstate distributions at the start and end times; or, at least, at the beginning and shortly *after* the end time. Indeed, the local distributions associated with each memory state are then nearly local-equilibrium distributions. If, to the contrary, the ending microstate distribution $\boldsymbol{\mu}_\tau(s)$ is *not* yet metastable, then it quickly *relaxes* to the metastable distribution $\boldsymbol{\mu}_{\tau + \delta t}(s) \approx \sum_{m \in \boldsymbol{\mathcal{M}}} \boldsymbol{\mu}_\tau(m) \boldsymbol{\pi}_{x_\tau}^{(m)}(s)$. This results in yet more dissipation.

To include all dissipation associated with a computation, we extend a given protocol's start and end times to include most of the post-computation relaxation to metastability. Assuming a metastable distribution at the beginning and end of the computation means $\boldsymbol{\mu}_0^{(m)} \approx \boldsymbol{\pi}_{x_0}^{(m)}$ and $\boldsymbol{\mu}_\tau^{(m)} \approx \boldsymbol{\pi}_{x_\tau}^{(m)}$ and the full microstate distributions are then $\boldsymbol{\mu}_0 \approx \sum_{m \in \boldsymbol{\mathcal{M}}} \boldsymbol{\mu}_0(m) \boldsymbol{\pi}_{x_0}^{(m)}$ and $\boldsymbol{\mu}_\tau \approx \sum_{m \in \boldsymbol{\mathcal{M}}} \boldsymbol{\mu}_\tau(m) \boldsymbol{\pi}_{x_\tau}^{(m)}$,

respectively. Thus, the beginning and ending microstate distributions are almost entirely determined by the distribution over memory states.

Initial and final memory-system metastability means that the empirically-observable memory transitions are always implicitly conditioned on metastability:

$$\Pr_{x_{0:\tau}}\left(\mathcal{M}_\tau = m' | \mathcal{M}_0 = m\right)$$
$$= \Pr_{x_{0:\tau}}\left(\mathcal{M}_\tau = m' | \mathcal{M}_0 = m, \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_0}^{(m)}\right) .$$

Similarly, if we operate the control protocol in reverse, starting metastably under the influence of $x_\tau^\dagger$, the observed memory-transition probabilities are:

$$\Pr_{\boldsymbol{\mathit{A}}(x_{0:\tau})}\left(\mathcal{M}_\tau = m^\dagger | \mathcal{M}_0 = m'^\dagger\right)$$
$$= \Pr_{\boldsymbol{\mathit{A}}(x_{0:\tau})}\left(\mathcal{M}_\tau = m^\dagger | \mathcal{M}_0 = m'^\dagger, \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_\tau^\dagger}^{(m'^\dagger)}\right) . \quad (12)$$

Note, for comparison with Eq. (11), that:[6]

$$\boldsymbol{\pi}_x^\dagger(s) = \boldsymbol{\pi}_x(s^\dagger) = \boldsymbol{\pi}_{x^\dagger}(s) .$$

And so, we have:

$$\boldsymbol{\pi}_{x_\tau^\dagger}^{(m'^\dagger)} = \boldsymbol{\pi}_{x_\tau}^{\dagger(m'^\dagger)} = \boldsymbol{\mu}_\tau^{\dagger(m'^\dagger)} .$$

This allows us to remove all dependence on microstates from the work bound and estimate dissipation purely from observed memory-state trajectories. In particular, Eq. (11) simplifies to:

$$d(m, m') = \Pr_{x_{0:\tau}}\left(\mathcal{M}_\tau = m' | \mathcal{M}_0 = m\right)$$
$$\times \ln \frac{\Pr_{x_{0:\tau}}(\mathcal{M}_\tau = m' | \mathcal{M}_0 = m)}{\Pr_{\boldsymbol{\mathit{A}}(x_{0:\tau})}(\mathcal{M}_\tau = m^\dagger | \mathcal{M}_0 = m'^\dagger)} . \quad (13)$$

Now, the dissipated work bound can be expressed in terms only of the probabilities of memory-state inputs and outputs:

$$\beta\langle W_{\text{diss}}\rangle \geq \Delta\,\text{H}(\mathcal{M}_t) + \sum_{m,m'} \Pr_{x_{0:\tau}}\left(\mathcal{M}_\tau = m', \mathcal{M}_0 = m\right)$$
$$\times \ln \frac{\Pr_{x_{0:\tau}}(\mathcal{M}_\tau = m' | \mathcal{M}_0 = m)}{\Pr_{\boldsymbol{\mathit{A}}(x_{0:\tau})}(\mathcal{M}_\tau = m^\dagger | \mathcal{M}_0 = m'^\dagger)} .$$

This bounds a computation's dissipated energy using only knowledge of the memory-state transitions that result from forward driving $x_{0:\tau}$ and reverse driving $\boldsymbol{\mathit{A}}(x_{0:\tau})$.

_____

[6] That $(\boldsymbol{\pi}_x)^\dagger = \boldsymbol{\pi}_{x^\dagger}$ follows from the energy eigenvalues being invariant under time reversal: $E_x(s) = E_{x^\dagger}(s^\dagger)$. Note that equilibrium probability depends only on the energy of the state and on the partition function. $E_x(s) = E_{x^\dagger}(s^\dagger)$ implies $Z_x = Z_{x^\dagger}$. These properties together therefore imply $(\boldsymbol{\pi}_x)^\dagger = \boldsymbol{\pi}_{x^\dagger}$.

However, additional simplifications arise, if the driving is time-symmetric.

### D. Time Symmetric Driving

Let us now consider the consequences for reliable computing with *time-symmetric* protocols—those for which $x_{0:\tau} = \boldsymbol{\mathit{A}}(x_{0:\tau})$—as shown in the upper panel of Fig. 2's computation. In these cases, Eq. (13) simplifies considerably:

$$d(m, m') = \Pr_{x_{0:\tau}}\left(\mathcal{M}_\tau = m' | \mathcal{M}_0 = m\right)$$
$$\times \ln\left(\frac{\Pr_{x_{0:\tau}}(\mathcal{M}_\tau = m' | \mathcal{M}_0 = m)}{\Pr_{x_{0:\tau}}(\mathcal{M}_\tau = m^\dagger | \mathcal{M}_0 = m'^\dagger)}\right) . \quad (14)$$

With Eqs. (10) and (14), we arrive at the remarkable result that the minimal dissipation depends only on memory-transition probabilities actually exercised by the computation:

$$d(m, m') = p(m \to m') \ln \frac{p(m \to m')}{p(m'^\dagger \to m^\dagger)} ,$$

where we rewrote the result suggestively denoting the computation as $p(m \to m') = \Pr_{x_{0:\tau}}(\mathcal{M}_\tau = m' | \mathcal{M}_0 = m)$. (Conveniently, no longer is there a dependence on counterfactual probabilities that *would* be induced by time-reversed driving.) Thus, the bound on dissipated work $\langle W_{\text{diss}}\rangle_{\text{min}}^{t-\text{sym}}$ includes a term $\Delta H(\mathcal{M})$ that depends on initial and final observable distributions and one that depends on the transition paths between memory states:

$$\beta\langle W_{\text{diss}}\rangle_{\text{min}}^{t\text{-sym}} = \Delta H(\mathcal{M}_t)$$
$$+ \sum_{m,m'} \boldsymbol{\mu}_0(m)p(m \to m') \ln \frac{p(m \to m')}{p(m'^\dagger \to m^\dagger)} . \quad (15)$$

This demonstrates how restricting to time-symmetric control leads to unavoidable dissipation that depends explicitly on a computation's logic.

This adds to a short list of work bounds determined by the computation implemented. Original on this list—the minimal work required for a computation without control restrictions—is simply Landauer's bound:

$$\langle W\rangle_{\text{min}}^{\text{Landauer}} \equiv -k_{\text{B}}T\Delta\,\text{H}(\mathcal{M}_t) ,$$

when memory states all have equal local free energies. More generally, though, as seen from Eq. (4), the minimal work required for a computation is:

$$\langle W\rangle_{\text{min}} \equiv \Delta\mathcal{F} \quad (16)$$
$$= \Delta\langle F_{x_t}^{(\mathcal{M}_t)}\rangle + \langle W\rangle_{\text{min}}^{\text{Landauer}} , \quad (17)$$

which is achieved in the limit of zero dissipation. Here, we recalled that for metastable distributions the nonequilibrium free energy decomposes as:

$$\mathcal{F} = \langle F_{x_t}^{(\mathcal{M}_t)} \rangle - k_B T \, \text{H}(\mathcal{M}_t) \, ,$$

where $F_x^{(m)} \equiv -k_B T \ln \left( \sum_{s \in m} e^{-\beta E_x(s)} \right)$ is the local-equilibrium free energy associated with the memory $m$ [4, 5].

An important related but often overlooked result emerges. Even for logically-irreversible computations, the minimal work required by this generalized Landauer bound is *reversible*: it corresponds to a stored change in nonequilibrium free energy that can be retrieved. Thus, ultimately no work needs to be dissipated. We *could* have $\langle W_{\text{diss}} \rangle = 0$ in the case of unrestricted control. And, this alerts us to additional opportunities for optimizing implementations. Appealing to Eq. (4) again, we see that the minimum total work under time-symmetric driving must be:[7]

$$\langle W \rangle_{\text{min}}^{t\text{-sym}} = \Delta \langle F_{x_t}^{(\mathcal{M}_t)} \rangle + \langle W \rangle_{\text{min}}^{\text{Landauer}} + \langle W_{\text{diss}} \rangle_{\text{min}}^{t\text{-sym}}$$
$$= \Delta \langle F_{x_t}^{(\mathcal{M}_t)} \rangle + k_B T \sum_{m,m'} \boldsymbol{\mu}_0(m) p(m \to m') \ln \frac{p(m \to m')}{p(m'^\dagger \to m^\dagger)} \, . \tag{18}$$

The consequences are immediate. Computing with time-symmetric protocols requires additional work above and beyond the Landauer bound. Moreover, all of this work—beside the reversible Landauer contribution and the change in local free energy—must be irreversibly dissipated since it is not stored as nonequilibrium free energy. $\langle W_{\text{diss}} \rangle > 0$, quantified precisely by Eq. (15), contributes to wasted heat in any computer. It adds on top of other beyond-Landauer work bounds including (i) the cost of modular computations, where global correlations are dissipated due to localized control [5, 33] and (ii) the cost of neglecting the local statistics of the manipulated memory [5, 34].

### E. (Non)Reciprocity

We argued that it is common—in practice and, occasionally, out of necessity—for a computation to be implemented by transforming metastable memories with time-symmetric driving. We are now in a position to plainly state the thermodynamic consequences of this type of computation.

Generically, a physical computation produces and is characterized by the set of memory transition probabilities: $\{ p(m \to m') \}_{m,m' \in \mathcal{M}}$. From Eq. (18), we see that the work required for such a computation, beyond the change in local free energy (which is often constructed to be zero),

is determined by the *nonreciprocity* of memory transitions. We say that a memory transition is *reciprocated* if:

$$p(m \to m') = p(m'^\dagger \to m^\dagger) \, . \tag{19}$$

*Nonreciprocity* $\Psi$ quantifies the deviation from this:

$$\Psi(m \to m') \equiv \ln \frac{p(m \to m')}{p(m'^\dagger \to m^\dagger)} \, . \tag{20}$$

Notably, it vanishes when Eq. (19) is satisfied. Keep in mind that $m^\dagger$ and $m'^\dagger$ are themselves valid memory states, so nonreciprocity compares the probability of two different memory transitions.

Appendix B establishes the transition-specific version of Eq. (18). To summarize, we first derive a useful transition-specific fluctuation theorem:

$$\langle e^{-\beta W} \rangle_{\text{Pr}_{x_{0:\tau}}(\mathcal{S}_{0:\tau} | \mathcal{M}_\tau = m', \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_0}^{(m)})}$$
$$= e^{-\beta \Delta F_x^{(\mathcal{M})}} \frac{\text{Pr}_{\boldsymbol{\mathcal{H}}(x_{0:\tau})}(\mathcal{M}_\tau = m^\dagger | \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_\tau^\dagger}^{(m'^\dagger)})}{\text{Pr}_{x_{0:\tau}}(\mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_0}^{(m)})}, \tag{21}$$

which applies to systems that start in local equilibrium. The appearance of the local-equilibrium free energies makes this a useful generalization of related past results [35]. By Jensen's inequality, and assuming time-symmetric control, this then implies that the minimal average work required of a memory transition is:

$$W_{\text{min}}^{t\text{-sym}}(m \to m') = F_{x_0}^{(m')} - F_{x_0}^{(m)} + k_B T \Psi(m \to m'). \tag{22}$$

In the common scenario where all local-equilibrium free

---

[7] Common computational substrates have $\Delta \langle F_{x_t}^{(m)} \rangle = 0$, since engineered memories typically have equal local-equilibrium energies. The change in local free energy is nevertheless likely important in the manipulation of biological memories.

energies are constructed to be equal, this reduces to:

$$W_{\min}^{t\text{-sym}}(m \to m') = k_B T \Psi(m \to m') . \qquad (23)$$

In other words, *nonreciprocity implies work.* This can also be seen from *Trajectory Class Fluctuation Theorems* if one restricts to trajectories that start and end in the specified memory states [36]. Not all work is bad, though. When $W_{\min}^{t\text{-sym}}(m \to m')$ is negative, some work can be harvested during the transition. In fact, such work harvesting is necessary to achieve minimal dissipation.

Equation (23) implies that *strictly reciprocal* computations, for which Eq. (19) is satisfied for all $m, m' \in \mathcal{M}$, can be implemented with a time-symmetric protocol with no expended work at all. The identity map is a trivial example. Two-cycles, like a bit swap, can also be implemented for free if using a time-symmetric memory substrate (for which $m^\dagger = m$ for all $m \in \mathcal{M}$). However, most computations require nonreciprocated memory transitions and, so, require work.

While the Landauer work cost corresponds to average state-space compression, nonreciprocity corresponds to a localized imbalance of memory currents. Even in the case of heterogeneous local free energies (i.e., Eq. (22)), the minimal dissipation required of time-symmetrically driven computations is the difference between the two:

$$\langle W_{\mathrm{diss}} \rangle_{\min}^{t\text{-sym}} = k_B T \left\langle \Psi(\mathcal{M}_0 \to \mathcal{M}_\tau) \right\rangle - \langle W \rangle_{\min}^{\mathrm{Landauer}}. \quad (24)$$

The expected nonreciprocity $\langle \Psi(\mathcal{M}_0 \to \mathcal{M}_\tau) \rangle$ weights how often a nonreciprocated transition is exercised. This uses the *weighted nonreciprocity* $d(m \to m')$:

$$d(m \to m') = p(m \to m')\Psi(m \to m') , \qquad (25)$$

which appeared previously in Eq. (14). The nonreciprocity weight is critical to the following analysis. Though $\Psi(m \to m') = -\Psi(m'^\dagger \to m^\dagger)$, weighted nonreciprocity can strongly break the symmetry.

Consider a typical computation that enforces logical transitions: some transitions should happen with certainty while others should be forbidden. In particular, consider a nonreciprocated $m \mapsto m'$ logical transition: $m \mapsto m'$ should be strongly enforced ($p(m \to m') \approx 1$), while $m'^\dagger \not\mapsto m^\dagger$ should be strongly forbidden ($p(m'^\dagger \to m^\dagger) < \epsilon \ll 1$ for some error tolerance $\epsilon \to 0$). Then $d(m \to m')$ diverges as $\ln(1/\epsilon)$ while $d(m'^\dagger \to m^\dagger)$ vanishes. In short, divergent work is required to implement the desired nonreciprocated $m \mapsto m'$ logical transition.

A *deterministic* computation $\mathcal{C}$ entails a deterministic logical transformation $m \mapsto \mathcal{C}(m)$ satisfying the reciprocity condition of Eq. (19) only when:

$$\mathcal{C}\big(\mathcal{C}(m)^\dagger\big)^\dagger = m . \qquad (26)$$

If Eq. (26) is *not* satisfied—i.e., if $\mathcal{C}(m) = m'$ but $\mathcal{C}(m'^\dagger) \neq m^\dagger$—then the logical transformation appears to require *infinite* work and *infinite* dissipation since, from Eq. (18), it appears to require work of $k_B T \ln \frac{1}{0}$. In practice, since dissipation is bounded, such logical transformations can only be approximated. Reliable logic (with a probability of error less than some small $\epsilon$) requires significant dissipation of at least $k_B T \ln(1/\epsilon)$ for each violation of Eq. (26). Section III develops the consequences of this new relationship between reliability and dissipation. That then allows us to explore how the two can be balanced in several examples.

### F. Time Symmetric Memory

The following explores the implications of these results for time-reversal-invariant memories—$m = m^\dagger$—in which information is stored via time-symmetric variables, such as spatial location or physical conformation. In these cases, Eq. (15) simplifies to:

$$\beta \langle W_{\mathrm{diss}} \rangle_{\min}^{t\text{-sym}} = \Delta H(\mathcal{M}_t)$$
$$+ \sum_{m,m'} \boldsymbol{\mu}_0(m) p(m \to m') \ln \frac{p(m \to m')}{p(m' \to m)} . \quad (27)$$

As before, minimal dissipation depends directly on the nonreciprocity between memory transitions, but now we have the simplification that:

$$\Psi(m \to m') = \ln \frac{p(m \to m')}{p(m' \to m)} .$$

And here, reciprocity has an especially straightforward interpretation: If $m$ maps to $m'$, then $m'$ should map back to $m$ with the same probability. Otherwise work is required and, if it is above the Landauer bound, it must be dissipated.

What is reciprocity in the limit of deterministic computation? For time-symmetric memory, it acts somewhat like logical invertibility. However, it is stricter. Rather, it is logical self-invertibility. Logical noninvertibility, in contrast, gives rise to state-space compression—the origin of the Landauer bound. Indeed, circumventing the latter led early researchers to investigate *reversible* computing—computing with invertible logical transformations [1]. Logically reversible computing requires no work to operate. However, it is now understood that even logically irreversible computing can be accomplished with zero dissipation, despite requiring some recyclable work [2–4].

The next section draws out the consequences of nonreciprocity in the limit of nearly deterministic computation. In contrast to the Landauer cost of logical irreversibility—which, in the limit of zero error, saturates to some small

value on the order of the thermal energy—the thermodynamic cost of nonreciprocity *diverges* in the limit of zero error, with no chance of recovering the input energy. It demands an accounting.

## III. ALMOST-DETERMINISTIC COMPUTING

At this point, we showed that the transition probabilities entailed by a desired logical transformation set a lower bound on the work required to instantiate it physically. We noted that this gives a much stronger bound than that due to Landauer, which only depends on the relative state-space volume supporting memory before and after a computation. More specifically, when time-symmetric control implements a transformation of metastable memories, the minimal work investment is proportional to the nonreciprocity (non-self-invertibility) of memory transitions. Furthermore, we discovered that the time-reversal symmetries of the memory elements themselves can substantially alter the minimal thermodynamic costs.

Drawing out the practical consequences, the following shows that computing with minimum work depends directly on a computation's error rate $\epsilon$. Most concretely, time-symmetry-induced dissipation diverges as $\ln(1/\epsilon)$ for almost-deterministic computation, where metastable memories are transformed at low error rate $\epsilon \to 0$.

Consider the physical memory apparatus of a computer that approximately implements a deterministic computation $\mathcal{C} \colon \boldsymbol{\mathcal{M}} \to \boldsymbol{\mathcal{M}}$, mapping memory states to memory states. Then, the memory-state transition probabilities appearing in Eq. (15) are strongly constrained by the computation's desired reliability. That is, the probability that the nonequilibrium thermodynamic information processing takes a memory state $m$ to anywhere besides $\mathcal{C}(m)$ must be no greater than the error tolerance $\epsilon$, with $0 < \epsilon \ll 1$. Reliability requires choosing a time-symmetric drive protocol $x_{0:\tau}$ that guarantees that $\mathrm{Pr}_{x_{0:\tau}}\big(\mathcal{M}_\tau = m' | \mathcal{M}_0 = m\big) \le \epsilon$ for all $m, m' \in \boldsymbol{\mathcal{M}}$ such that $m' \ne \mathcal{C}(m)$. The thermodynamic implications for a reliable computation then follow from Eq. (15) with:

$$p(m \to m') \begin{cases} \ge 1 - \epsilon & \text{if } \mathcal{C}(m) = m' \\ \le \epsilon & \text{if } \mathcal{C}(m) \ne m' \end{cases} .$$

Evaluating Eq. (15) requires addressing four cases, depending on whether $\mathcal{C}(m) = m'$ or $\mathcal{C}(m) \ne m'$ and on whether $\mathcal{C}(m'^\dagger) = m^\dagger$ or $\mathcal{C}(m'^\dagger) \ne m^\dagger$. Any given implementation results in an *actual* probability of error for each of the intended transitions: $\epsilon_m = 1 - p\big(m \to \mathcal{C}(m)\big)$. We adopt the design constraint that $\epsilon_m \le \epsilon$ for all possible initial memories $m$. Then, given an implementation $x_{0:\tau}$, the probability of an *accidental* memory transition is $\epsilon_{m\to m'} = p(m \to m')$ for $m' \ne \mathcal{C}(m)$. Since

$\sum_{m' \in \boldsymbol{\mathcal{M}} \backslash \{\mathcal{C}(m)\}} \epsilon_{m \to m'} = \epsilon_m \le \epsilon$, we must have that $0 \le \epsilon_{m \to m'} \le \epsilon_m \le \epsilon$.

To simplify, we temporarily restrict to the case of time-reversible memories, where $m = m^\dagger$. The four cases then depend on whether $\mathcal{C}(m) = m'$ or $\mathcal{C}(m) \ne m'$, and on whether $\mathcal{C}(m') = m$ or $\mathcal{C}(m') \ne m$. Later, we quote the general result that does not require $m = m^\dagger$.

Appendix C uses the error constraints to evaluate the weighted nonreciprocity:

$$d(m, m') = p(m \to m') \ln \frac{p(m \to m')}{p(m' \to m)} \ ,$$

for the four possible cases, finding:

1. $\mathcal{C}(m) = m'; \mathcal{C}(m') = m$:
   $$-\epsilon \le d^{(1)}(m, m') \le \epsilon + \tfrac{1}{2}\epsilon^2 + \mathcal{O}(\epsilon^3) \ .$$
2. $\mathcal{C}(m) = m'; \mathcal{C}(m') \ne m$:
   $$\ln(\epsilon^{-1}) \lesssim d^{(2)}(m, m') \le \ln(\epsilon_{m' \to m}^{-1}) \ .$$
3. $\mathcal{C}(m) \ne m'; \mathcal{C}(m') = m$:
   $$\epsilon_{m \to m'} \ln \epsilon_{m \to m'} < d^{(3)}(m, m') < 0 \ .$$
4. $\mathcal{C}(m) \ne m'; \mathcal{C}(m') \ne m$:
   $$-\epsilon/e \le d^{(4)}(m, m') = \epsilon_{m \to m'} \ln \left( \frac{\epsilon_{m \to m'}}{\epsilon_{m' \to m}} \right) \ .$$

In the high-reliability limit $\epsilon \to 0$, most $d^{(n)}(m, m')$ terms are on the order of $\epsilon$ and so tend to zero. That is, *except* for $d^{(2)}(m, m')$ in Case 2 whose contribution to the dissipation *diverges* as $\epsilon \to 0$ since $d^{(2)}(m, m') \gtrsim \ln(\epsilon^{-1})$.

Case 4 is somewhat more delicate than Cases 1 and 3, due to the ratio of errors. When $\epsilon_{m' \to m} \ge \epsilon_{m \to m'}$, then $-\epsilon/e \le d^{(4)}(m, m') \le 0$, so that $d^{(4)}(m, m')$ vanishes as $\epsilon \to 0$. However, when $\epsilon_{m \to m'} > \epsilon_{m' \to m}$, there is a chance for slightly more dissipation. Nevertheless, $d^{(4)}(m, m')$ is still only on the order of $\epsilon$, so long as the relative error rates $\epsilon_{m \to m'}$ and $\epsilon_{m' \to m}$ are within several orders of magnitude of each other. If $\epsilon_{m \to m'} \gg \epsilon_{m' \to m}$, then extra dissipation will be incurred due to the stringent reliability of the forbidden $m \nrightarrow m'$ transition, beyond the design constraint.

The upper bound in Case 2 also implies extra dissipation when $\epsilon_{m' \to m}$ is more reliable than the reliability design constraint. And, this extra dissipation need not be small. Generically, though, for a given error tolerance, we expect that minimal dissipation can be achieved by allowing all error rates to be as close to uniform as possible, meeting but not significantly exceeding the overall reliability constraint. Still, due to Case 2, even the minimal dissipation diverges with increasing reliability.

Thus, Case 2 provides the only contribution of the $d(m, m')$ terms to Eq. (10) in the low-error limit. Since the target computation is deterministic—$\mathcal{C}$ maps each memory state $m$ to one memory state $\mathcal{C}(m)$—there can be at most one contribution of $d^{(2)}(m, m')$ for each $m$,

coming from $m' = \mathcal{C}(m)$, and only if $\mathcal{C}(m') \neq m$. The total contribution from each $m$ is:

$$\big[\mathcal{C}(\mathcal{C}(m)) \neq m\big] \ln(\epsilon^{-1}_{\mathcal{C}(m) \to m}) \ ,$$

where $\big[\cdot\big]$ is the Iverson bracket, which returns 1 when its argument is true and 0 otherwise. In this case, $\big[\mathcal{C}(\mathcal{C}(m)) \neq m\big] = 1 - \delta_{m,\mathcal{C}(\mathcal{C}(m))}$.

Applying the low-$\epsilon$ contributions to Eq. (14) yields an approximate bound for the time-symmetric work. If the memories have the same local-equilibrium free energy, the bound is:

$$
\begin{aligned}
\beta \langle W \rangle^{\text{t-sym}}_{\min} &\approx \sum_{m \in \mathcal{M}} \boldsymbol{\mu}_0(m) \big[\mathcal{C}(\mathcal{C}(m)) \neq m\big] \ln(\epsilon^{-1}_{\mathcal{C}(m) \to m}) \\
&\geq \ln(\epsilon^{-1}) \sum_{m \in \mathcal{M}} \boldsymbol{\mu}_0(m) \big[\mathcal{C}(\mathcal{C}(m)) \neq m\big] \quad (28) \\
&= \beta \langle W \rangle^{\text{approx}}_{\min} \ .
\end{aligned}
$$

This determines how the tradeoff between work and reliability scales for almost-perfect computation. And, it identifies the role that a computation's reciprocity plays in dissipation.

The result is a simple error–dissipation tradeoff, if we restrict ourselves to using time-reversible protocols for implementing reliable computations:

$$
\begin{aligned}
\beta \langle W_{\text{diss}} \rangle^{\text{t-sym}}_{\min} &\gtrsim \beta \langle W_{\text{diss}} \rangle^{\text{approx}}_{\min} \\
&= \big\langle \big[\mathcal{C}(\mathcal{C}(\mathcal{M}_0)) \neq \mathcal{M}_0\big]\big\rangle_{\mathcal{M}_0} \ln(\epsilon^{-1}) \\
&\quad + \Delta H(\mathcal{M}_t) \ . \quad (29)
\end{aligned}
$$

Since the change $\Delta \mathrm{H}(\mathcal{M}_t)$ in entropy is finite as $\epsilon \to 0$, the divergent nonreciprocity contribution to the dissipation dominates. Moreover, the simple scaling of the reliability–dissipation tradeoff is the same as that for the reliability–work tradeoff. In the low-error limit, the minimal nonreciprocity depends only on the computation:

$$\langle \Psi(\mathcal{M}_0 \to \mathcal{M}_\tau) \rangle \geq \big\langle \big[\mathcal{C}(\mathcal{C}(\mathcal{M}_0)) \neq \mathcal{M}_0\big]\big\rangle_{\mathcal{M}_0} \ln(\epsilon^{-1}).$$

The reciprocity coefficient:

$$\big\langle \big[\mathcal{C}(\mathcal{C}(\mathcal{M}_0)) \neq \mathcal{M}_0\big]\big\rangle_{\mathcal{M}_0} = \sum_{m \in \mathcal{M}} \boldsymbol{\mu}_0(m) \big[\mathcal{C}(\mathcal{C}(m)) \neq m\big]$$

is the probability that the memory makes a transition that is *not* reciprocated if the output becomes the input to the computation. In short, *computations with nonreciprocity require significant dissipation when implemented with a time-symmetric protocol.*

A sequel explores the dynamical mechanism for the dissipation in this case. With time-symmetric protocols, state-space compression occurs only via a loss of stability and subsequent dissipation. This corresponds to a topological restriction on the bifurcation structure of lo-

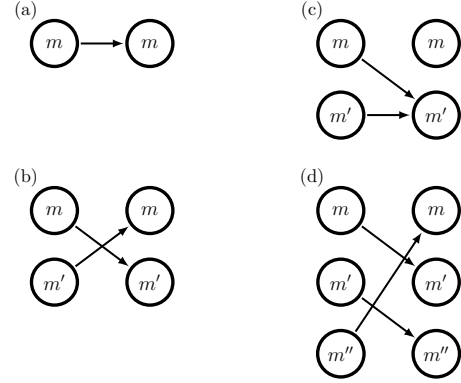cal attractors—a restriction imposed by time-symmetric control.



FIG. 4. Simple features of deterministic computations determine the minimal dissipation required to implement them reliably with time-symmetric protocols. Dissipation is incurred for nonreciprocated memory transitions. When using time-reversal-invariant memory elements ($m^\dagger = m$): (a) identity mappings and (b) swap operations can be implemented freely, whereas (c) memory-space compression and (d) nonreciprocated permutations require dissipation that diverges with increasing reliability.

One also concludes that strictly-reciprocal logic gates, such as the identity (communication relay) and NOT gates, where $\mathcal{C}(\mathcal{C}(m))$ always returns $m$, are *exempt* from this dissipation. One can efficiently implement them with time-symmetric protocols. For all other computations, time-*asymmetric* control must be used to avoid this dissipation. Figure 4 gives examples of reciprocated and nonreciprocated memory transitions.

The main result Eq. (28) generalizes to the case where $m, m^\dagger \in \mathcal{M}$ without the requirement that $m = m^\dagger$. Then, dissipation diverges with increasing reliability whenever $m \to m'$ transitions are made such that $\mathcal{C}(m) = m'$ but $\mathcal{C}(m'^\dagger) \neq m^\dagger$. For uniform local free energies, Eq. (28) becomes:

$$\beta \langle W \rangle^{\text{approx}}_{\min} = \ln(\epsilon^{-1}) \sum_{m \in \mathcal{M}} \boldsymbol{\mu}_0(m) \big[\mathcal{C}(\mathcal{C}(m)^\dagger) \neq m^\dagger\big]. \quad (30)$$

Notably, this allows one to quantify error–work and error–dissipation tradeoffs when computing with magnetic memory systems, for example. For a single bistable magnetic device, $m$ could represent the 1 memory of having an "up" magnetic moment or "clockwise" in the case of toroidal magnetic core memory elements. And, $m^\dagger$ would then represent the "down" or "counter-clockwise" memory 0, as shown on the right side of Fig. 3. This implies that memories with certain types of symmetry are better suited for certain types of logical operation, in that they minimize nonreciprocity and so dissipation.

To emphasize, the dissipations here are distinct from the

heat associated with logically-irreversible transformations, as discussed by Landauer and Bennett [1, 19], which arises as a compensation to microscopic state-space contraction. There are two important distinctions.

First, the heat devolved with logical irreversibility—e.g., the minimal heat of $k_{\mathrm{B}}T \ln 2$ released to the environment upon erasure—is *not necessarily irreversibly dissipated* [2–5]. It is offset by a change in the nonequilibrium addition to free energy and so can be leveraged later to do an equal amount of useful work. In contrast, nonreciprocity dissipation is energy that is truly dissipated. It is irretrievably lost to the environment, with no compensation via change of the nonequilibrium addition to free energy and so can never be recovered.

Second, logical nonreciprocity is distinct from logical irreversibility. Strictly-reciprocal computations, where $\mathcal{C}(\mathcal{C}(m)) = m$ for all $m \in \boldsymbol{\mathcal{M}}$, are logically reversible, being their own inverses. However, logically reversible permutations of the memory can be completely nonreciprocal, as demonstrated in Fig. 4(d). Reciprocity requires not only that the deterministic logic be invertible, but further that the logical dynamic inverts itself. In short, the Landauer work cost corresponds to logical noninvertibility, while nonreciprocity cost corresponds to logical *noninvolution*. The reciprocity bound on dissipation can therefore be interpreted as the minimal work required to implement a reliable computation with time-symmetric protocols *in addition to* the well-known Landauer bound.

## IV.  APPLICATIONS

With the basics of time-symmetric and nonreciprocal computing in hand, we turn to explore the thermodynamic implications for erasure, logic gates, and biological information processing. Our companion work Ref. [37] gives a detailed analysis of erasure and explores several model implementations.

### A.  Erasure

To directly illustrate how the cost of time-symmetric control differs from Landauer's bound, consider the classic example of bit erasure. Landauer originally described a method for erasing a bit of information stored in a double-well potential in contact with a heat bath: tilt the potential—moving overdamped stochastic particles to one side—and return the potential to its original orientation, leaving the particles (temporarily) trapped in one well [19]. If implemented naively, via a protocol that raises the energy of one well and then lowers it at the same rate, the time-symmetry results in dissipation significantly above Landauer's bound, as described by the bounds just

developed. Note that protocol time-symmetry does *not* imply spatial symmetry—one can tilt the potential to the left without needing to tilt it to the right.

Bit erasure $\mathcal{C}_{\mathrm{erase}}$ operates on a single bistable element with memory states $\boldsymbol{\mathcal{M}} = \{\mathrm{L}, \mathrm{R}\}$ that correspond to occupying the Left or Right side of a double-well potential energy landscape. The computation is defined by resetting to the L memory state: $\mathcal{C}_{\mathrm{erase}}(\mathrm{L}) = \mathrm{L}$ and $\mathcal{C}_{\mathrm{erase}}(\mathrm{R}) = \mathrm{L}$. Hence, $\mathcal{C}_{\mathrm{erase}}(\mathcal{C}_{\mathrm{erase}}(\mathrm{L})) = \mathrm{L}$, while $\mathcal{C}_{\mathrm{erase}}(\mathcal{C}_{\mathrm{erase}}(\mathrm{R})) \neq \mathrm{R}$. Thus, the net transition probabilities $p(m \to m')$ shown in Fig. 5 characterize any reliable implementation of this computation in terms of the probabilities of errors, $\epsilon_{\mathrm{R}} = \epsilon_{\mathrm{R} \to \mathrm{R}}$ and $\epsilon_{\mathrm{L}} = \epsilon_{\mathrm{L} \to \mathrm{R}}$, that leave the system in the R state.
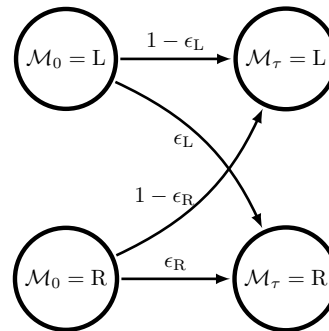


FIG. 5. Markov transition matrix for any implementation of erasure gives the error rate $\epsilon_{\mathrm{L}} = \epsilon_{\mathrm{L} \to \mathrm{R}}$ from initial memory state $\mathcal{M}_0 = \mathrm{L}$ and the error rate $\epsilon_{\mathrm{R}} = \epsilon_{\mathrm{R} \to \mathrm{L}}$ from the initial memory state $\mathcal{M}_0 = \mathrm{R}$.

From this, the only terms contributing to the bound on time-symmetric work are:

$$d(\mathrm{L}, \mathrm{R}) = \epsilon_{\mathrm{L}} \ln \frac{\epsilon_{\mathrm{L}}}{1 - \epsilon_{\mathrm{R}}} \text{ and}$$
$$d(\mathrm{R}, \mathrm{L}) = (1 - \epsilon_{\mathrm{R}}) \ln \frac{1 - \epsilon_{\mathrm{R}}}{\epsilon_{\mathrm{L}}} \ .$$

Allowing for a potentially nonuniform input distribution $\boldsymbol{\mu}_0 = (1 - p_{\mathrm{R}}, p_{\mathrm{R}})$ over memory states yields the exact bound on time-symmetric work investment:

$$\begin{aligned} \beta \langle W \rangle_{\min}^{t\text{-sym}} &= (1 - p_{\mathrm{R}}) d(\mathrm{L}, \mathrm{R}) + p_{\mathrm{R}} d(\mathrm{R}, \mathrm{L}) \\ &= (p_{\mathrm{R}}(1 - \epsilon_{\mathrm{R}}) - (1 - p_{\mathrm{R}})\epsilon_{\mathrm{L}}) \ln \frac{1 - \epsilon_{\mathrm{R}}}{\epsilon_{\mathrm{L}}} \\ &= (p_{\mathrm{R}} - \langle \epsilon \rangle) \ln \frac{1 - \epsilon_{\mathrm{R}}}{\epsilon_{\mathrm{L}}} \ , \end{aligned} \tag{31}$$

where $\langle \epsilon \rangle = p_{\mathrm{R}} \epsilon_{\mathrm{R}} + (1 - p_{\mathrm{R}}) \epsilon_{\mathrm{L}}$ is the average error of the computation.

Figure 6 plots the work requirements for an initially unbiased memory state with $p_{\mathrm{R}} = 1/2$. We see that the work diverges for small values of error $\epsilon_{\mathrm{L}}$. The plot also indicates that the work diverges for high values of error $\epsilon_{\mathrm{R}}$, due to the symmetry of the system between left and
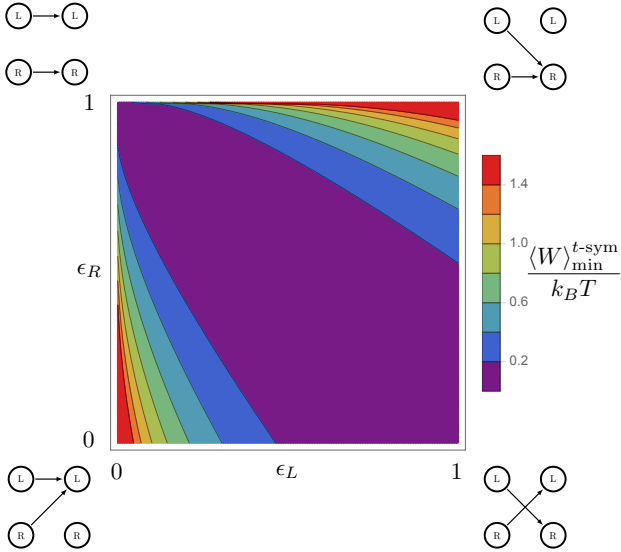
FIG. 6. Average minimum work (units of $k_B T$) required for a 1-bit operation diverges for low $\epsilon_L$ or high $\epsilon_R$, corresponding to erasure to L or erasure to R, respectively. The state-transition diagrams at the corners show the basic computations associated with each extremal point in computation parameter space. The identity map (upper left), the bit-flip (lower right), and every operation interpolated between require no work according to the time-symmetric control bound. The required work diverges at perfect erasure to Left (bottom-left corner) and perfect erasure to Right (top-right corner), where the red coloring covers all work values from 1.4 to $\infty$ $k_B T$.
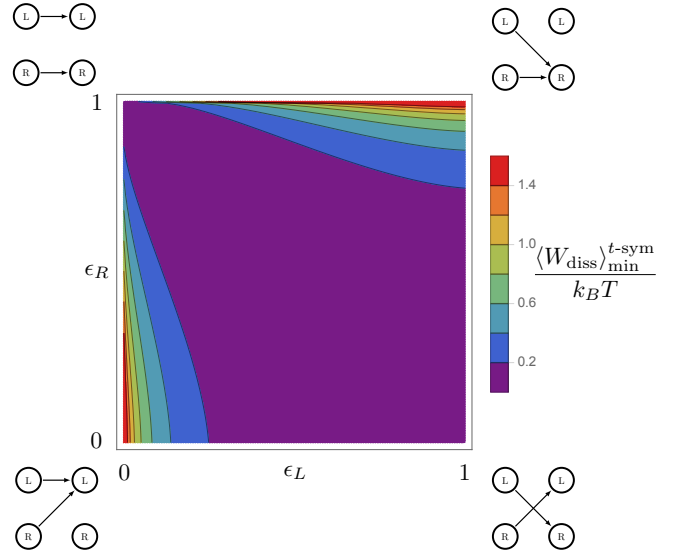


FIG. 7. Dissipated work—that required beyond the Landauer bound—behaves much the same as the minimum work over possible bit computations, diverging for high values of $\epsilon_R$ and low values of $\epsilon_L$ and vanishing for the identity map and bit flip. Dissipation diverges at the bottom-left and top-right corners, where the red coloring covers all values of dissipated work from 1.4 to $\infty$ $k_B T$.

right. It is also worth noting from the plot that there are computations, such as bit flips, which this bound suggests may be achievable with time-symmetric control and without energetic cost. However, we are primarily concerned with the lower left portion of the plot, where effective erasures occur. The divergent scaling of the work required to reliably erase overwhelms the rather meager energy requirements given by Landauer's bound. And, unlike the latter, this excess work requirement must be irreversibly dissipated. Subtracting off the Landauer bound $k_B T \Delta H(\mathcal{M}_t)$ to calculate the minimum dissipation $\langle W_{\text{diss}} \rangle_{\min}^{t\text{-sym}}$, Fig. 7 shows that the divergent contribution to the work is attributed to nonreciprocity.

Applying Eq. (28)'s work bounds to erasure, we see that:

$$\left\langle \llbracket \mathcal{C}_{\text{erase}}(\mathcal{C}_{\text{erase}}(\mathcal{M}_0)) \neq \mathcal{M}_0 \rrbracket \right\rangle_{\mathcal{M}_0} = p_R \ .$$

This means that, in the low-error limit:

$$\langle W \rangle_{\min}^{\text{approx}} = p_R \ln(\epsilon^{-1}) k_B T \ . \tag{32}$$

The initial memory-state entropy is $H_b(p_R) \equiv -p_R \ln p_R - (1-p_R)\ln(1-p_R) \leq \ln 2$, which is called the *binary entropy function* [24]. Whereas, the final memory-state entropy vanishes as $\epsilon \to 0$. Hence, from Eq. (29), we immediately find that the small-error dissipation necessary for time-

symmetric erasure diverges as $\epsilon \to 0$:

$$\langle W_{\text{diss}} \rangle_{\min}^{\text{approx}} \approx p_R \ln(\epsilon^{-1}) k_B T - H_b(p_R) k_B T \ . \tag{33}$$

Landauer erasure typically assumes a uniform initial distribution where $p_R = 1/2$. This results in $\langle W_{\text{diss}} \rangle^{t\text{-sym}} \gtrsim \frac{1}{2} \ln(1/4\epsilon) k_B T$. Since the memory-state entropy is bounded by $k_B T \ln 2$, the contribution from Landauer's bound is negligible compared to the $\ln(\epsilon^{-1})$ term for small $\epsilon$. And so, the latter dominates both the work and dissipation for high-fidelity erasure.

To test the strength of these bounds and their approximations, Ref. [37] considers two different implementations of time-symmetric erasure—rate equations for a two-level system and one-dimensional Langevin dynamics in a controlled double-well potential. The example implementations not only validate our bounds, but moreover show that the bounds are tight.

## B. Logic Gates

We next consider the minimal dissipation required to implement conventional two-input one-output logic gates that serve as the basis for modern digital computing: AND, NAND, OR, NOR, XOR, and the like. The typical implementation requires two bistable memory elements ($\mathcal{M}_t^{(\text{in}_1)}$ and $\mathcal{M}_t^{(\text{in}_2)}$) for the input and another bistable memory element ($\mathcal{M}_t^{(\text{out})}$) to robustly store the output.

Each memory element can take on one of two memory states $\mathcal{M}_t^{(\cdot)} \in \{0, 1\}$. Each logic gate is a computation $\mathcal{C}$ on the set of composite memory states $\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{M}}^{(\text{in}_1)} \times \boldsymbol{\mathcal{M}}^{(\text{in}_2)} \times \boldsymbol{\mathcal{M}}^{(\text{out})} = \{000, 001, 010, \dots 111\}$.

In particular, since networks of NAND gates are sufficient for universal computation, the NAND gate is worthy of immediate investigation. Appendix D analyzes the NAND gate, showing that: *if the memory elements are initiated statistically independent of each other*, then the minimal time-symmetric-implementation dissipation is comparable to that expected for bit erasure of the output bit, depending on how often the output bit is overwritten. However, when the memory system is initialized with correlation among the memory elements (perhaps as the correlated output of previous upstream computations), then the change in memory entropy is $\Delta \mathrm{H}(\mathcal{M}_t) \approx - \mathrm{H}\big(\mathcal{M}_0^{(\text{out})} \big| \mathcal{M}_0^{(\text{in}_1)}, \mathcal{M}_0^{(\text{in}_2)}\big)$ in the small $\epsilon$ limit, and the reciprocity coefficient is:

$$\Big\langle \big[\!\big[ \mathcal{C}_{\text{NAND}}(\mathcal{C}_{\text{NAND}}(\mathcal{M}_0)) \neq \mathcal{M}_0 \big]\!\big] \Big\rangle_{\mathcal{M}_0}$$
$$= \boldsymbol{\mu}_0(000) + \boldsymbol{\mu}_0(010) + \boldsymbol{\mu}_0(100) + \boldsymbol{\mu}_0(111) \ . \quad (34)$$

As a consequence, the initial correlation among the memory elements has a profound impact on the coefficient of minimal dissipation. The overall dissipation still diverges $\sim \ln(\epsilon^{-1})$, though, with increasing reliability as long as the implementation is time-symmetric.

This example highlights a previously unsuspected source of unnecessary and preventable dissipation in digital computers. Indeed, computations are currently implemented in hardware, whether via periodic clocking or static gate voltages, through time-symmetric driving. Error rates in today's reliable digital computers are extremely low. The so-called *soft error rate* (also known as *single-event upsets*) for 50 nm gate technology has been estimated to be on the order of $10^{-4}$ errors per $10^9$ hours of operation of a CMOS logic circuit [38]. Moreover, these errors are overwhelmingly due to cosmic rays rather than conventional thermal fluctuations. If the computer is properly shielded, the error rate is significantly lower. Using the aforementioned numbers for a conservative estimate and assuming a 3 GHz clock rate, this extreme reliability of $\epsilon \lesssim 10^{-26}$ errors/instruction would imply at least 30 $k_{\text{B}}T$ is dissipated in nearly every elementary operation in modern circuitry directly due to time-symmetric implementation. This is well beyond the Landauer bound that, for NAND and any other traditional two-input–one-output logic gate, is given by $-\Delta H(\mathcal{M}_t) \approx \mathrm{H}\big(\mathcal{M}_0^{(\text{out})} \big| \mathcal{M}_0^{(\text{in}_1)}, \mathcal{M}_0^{(\text{in}_2)}\big) \leq \ln 2$.

By way of contrast, the universal Fredkin gate—swap the state of the first two memory elements only if the third memory element is in state 1—is not only logically reversible, but is also a strictly-reciprocal gate [39]. As a strictly-reciprocal gate, it can implement univer-

sal computation efficiently using time-symmetric control. This makes Fredkin gates an especially tempting basis for energetically-efficient future computing technologies, assuming they can be made to transform metastable memories without the exponential sensitivity to initial conditions that condemned their initially-proposed mechanical instantiation [1, 40].

## C. Dissipation in Biological Processing

Beyond logical computation, the reliability-energy-efficiency tradeoff also applies to information processing in far-from-equilibrium biological systems. For this, it is useful to reformulate Eq. (3) to apply when $W_{\text{diss}}$ is the work dissipated in controlling (through $x_{0:\tau}$) a system already maintained out of equilibrium by a given housekeeping entropy production $\Sigma_{\text{hk}}$. This can happen, for example, via ATP hydrolysis or the maintenance of a chemical potential gradient. Then, according to Eq. (27) of Ref. [41], the more general starting point is:

$$\beta W_{\text{diss}} + k_{\text{B}}^{-1} \Sigma_{\text{hk}}$$
$$= \ln \left( \frac{\Pr_{x_{0:\tau}}\big(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{S}_0 \sim \boldsymbol{\mu}_0\big)}{\Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})}\big(\mathcal{S}_{0:\tau} = \boldsymbol{\mathcal{A}}(s_{0:\tau}) | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger\big)} \right) \ . \quad (35)$$

In steady-state biological processes, there is no external control. The controllable parameters can thus be considered as held constant, such that the driving protocol is trivially time-symmetric $\boldsymbol{\mathcal{A}}(x_{0:\tau}) = x_{0:\tau} = x_0 x_0 \dots x_0$ and $W_{\text{diss}} = 0$. That is, the system is driven solely by the housekeeping entropy production—power supplied from a stationary stochastic influence—that maintains the steady state. For example, molecular machines can sustain time-asymmetric functionality within their time-symmetric ambient chemical environment[8] due to the entropy they produce during chemical catalysis [15, 18]. In these nonequilibrium steady states, the expected microstate distribution of the system is the time-invariant steady-state distribution:

$$\boldsymbol{\mu}_0 = \boldsymbol{\mu}_\tau = \boldsymbol{\pi}_{\text{s.s.}} \quad (36)$$

Moreover, due to the low Reynolds numbers of the microbiological realm, momentum is heavily damped and the relevant memory states are time-reversal-invariant: $m^\dagger = m$ for all $m \in \boldsymbol{\mathcal{M}}$. We also have $\boldsymbol{\pi}_{\text{s.s.}}^\dagger = \boldsymbol{\pi}_{\text{s.s.}}$.

Integrating these generic features of steady-state biological

---

[8] This time-symmetric ambient chemical environment is maintained by power-consuming homeostatic mechanisms elsewhere in a cell.

transformations, Eq. (35) simplifies to:

$$k_{\mathrm{B}}^{-1}\Sigma_{\mathrm{hk}}=\ln\left(\frac{\Pr\left(\mathcal{S}_{0:\tau}=s_{0:\tau}|\mathcal{S}_0=s_0\right)\boldsymbol{\pi}_{\mathrm{s.s.}}(s_0)}{\Pr\left(\mathcal{S}_{0:\tau}=\boldsymbol{\mathcal{A}}(s_{0:\tau})|\mathcal{S}_0=s_\tau\right)\boldsymbol{\pi}_{\mathrm{s.s.}}(s_\tau)}\right). \quad (37)$$

The single-timestep version of this is well known:

$$\delta\Sigma_{\mathrm{hk}}=k_{\mathrm{B}}\ln\left(\frac{\Pr\left(\mathcal{S}_{\delta t}=s_{\delta t}|\mathcal{S}_0=s_0\right)\boldsymbol{\pi}_{\mathrm{s.s.}}(s_0)}{\Pr\left(\mathcal{S}_{0:\tau}=s_0|\mathcal{S}_0=s_{\delta t}\right)\boldsymbol{\pi}_{\mathrm{s.s.}}(s_{\delta t})}\right).$$

As $\delta t \to 0$, the relative probabilities reduce to the relative transition rates $\{r_{s\to s'}\}$:

$$\delta\Sigma_{\mathrm{hk}}=k_{\mathrm{B}}\ln\left(\frac{\boldsymbol{\pi}_{\mathrm{s.s.}}(s_0)\,r_{s_0\to s_{\delta t}}}{\boldsymbol{\pi}_{\mathrm{s.s.}}(s_{\delta t})\,r_{s_{\delta t}\to s_0}}\right). \quad (38)$$

(Metastability of mesoscopic conformations allows Eq. (38) to hold approximately for transitions between mesoscopic states as well.) For example, Eq. (38) was the starting point for the recently-derived thermodynamic uncertainty relations that received attention [42].

However, much can be learned from approaching finite-duration transformations directly, by coarse-graining Eq. (37) according to initial and final functionally-relevant states. Notably, it allows simple yet powerful analysis of the minimal dissipation required for biological functionality, regardless of how complicated the finite-duration biological implementations appear. And, it implies a generic error–dissipation tradeoff that we expect to apply broadly to the reliable performance of microbiological systems.

Working this out explicitly leads to a pleasantly simple picture of steady-state biological information processing and its requisite dissipation:

$$
\begin{aligned}
k_{\mathrm{B}}^{-1}\langle\Sigma_{\mathrm{hk}}\rangle &\geq \mathrm{D}_{\mathrm{KL}}\left[\Pr(\mathcal{M}_0=m,\mathcal{M}_\tau=m')\,||\,\Pr(\mathcal{M}_\tau=m',\mathcal{M}_{2\tau}=m)\right]\\
&= \underbrace{\Delta\,\mathrm{H}(\mathcal{M}_t)}_{=\,0\text{ since }\boldsymbol{\mu}_\tau=\boldsymbol{\mu}_0=\boldsymbol{\pi}_{\mathrm{s.s.}}} + \sum_{m,m'\in\boldsymbol{\mathcal{M}}}\boldsymbol{\pi}_{\mathrm{s.s.}}(m)\,p(m\to m')\ln\frac{p(m\to m')}{p(m'\to m)}\\
&= \sum_{m,m'\in\boldsymbol{\mathcal{M}}}\boldsymbol{\pi}_{\mathrm{s.s.}}(m)\,p(m\to m')\Psi(m\to m')\\
&= \langle\Psi(\mathcal{M}\to\mathcal{M}')\rangle_{\boldsymbol{\pi}_{\mathrm{s.s.}}(\mathcal{M})\,p(\mathcal{M}\to\mathcal{M}')}\,.
\end{aligned}
\quad (39)
$$

Steady-state computation quite simply becomes enforcing a Markov model for memory transitions $\{p(m\to m')\}$. The minimal dissipation achievable via any biological implementation then equals or exceeds the expected nonreciprocity. It should be noted that the steady-state distribution over memory states $\boldsymbol{\pi}_{\mathrm{s.s.}}$ is easily found as the stationary eigenstate of the Markov model's transition matrix.

In equilibrium, where $\boldsymbol{\pi}_{\mathrm{s.s.}}=\boldsymbol{\pi}_{\mathrm{eq}}$, the average nonreciprocity vanishes since detailed balance demands $\boldsymbol{\pi}_{\mathrm{eq}}(m)p_{\mathrm{eq}}(m\to m')=\boldsymbol{\pi}_{\mathrm{eq}}(m')p_{\mathrm{eq}}(m'\to m)$ while, as always, $\Psi(m\to m')=-\Psi(m'\to m)$. As necessary, this allows zero dissipation in equilibrium. In contrast, nonequilibrium steady-state computations must dissipate since they bias nonreciprocated transitions. These computations enforce nondetailed-balanced dynamics.

Equation (39) shows that dissipation in steady-state biological processes is bounded by the average nonreciprocity. Remarkably, this bound depends only on the net transformation and is independent of intermediary steps and biochemical details.

Thus, as in the preceding analysis, one expects a generic error–dissipation tradeoff:

$$d\Sigma_{\mathrm{hk}}/dt \propto \ln(1/\epsilon)$$

for reliable steady-state biological transformations of genetic information—whenever certain transitions need to be strongly enforced with only a small probability of failure $\epsilon$. The generic error–dissipation tradeoff holds, as an important example, for processes such as chemical proofreading necessary for reliable DNA replication. In this way, our contribution dovetails with more model-specific results on the latter by Refs. [43–46]. Since codons are nonreciprocally corrected to satisfy Watson–Crick base pairing, each correction requires significant dissipation.

The $d\Sigma_{\mathrm{hk}}/dt \propto \ln(1/\epsilon)$ error–dissipation tradeoff should arise in other reliable biological transformations as well. For example, to reliably maintain a target nonequilibrium state Ref. [47] finds exactly this minimum dissipation rate. This is relevant to protein folding, as well as other biological functionalities.

Our results also support and update the perspective of Ref. [48], where steady-state biochemical copying networks (e.g., for sensing) were found to have an unavoid-

able error–dissipation tradeoff due to time-invariant driving. Our results more broadly suggest error–dissipation tradeoffs for any biochemical task with nonreciprocity. It should now be clear that autonomous biochemical systems in steady state can never reach the Landauer bound on efficiency due to the time-symmetric control penalty.

## V. RELATED RESULTS

Let's consider how the tradeoffs between accuracy and energetic-efficiency under time-symmetric protocols compare to related approaches to thermodynamic control, both historically and currently.

In Ref. [49], Brillouin defines the reliability as $1/\epsilon$, where $\epsilon$ is the probability of reading out the wrong value due to thermal agitation. He found that the entropy production necessary for some methods of reliable *readout* is $k_B \ln(1/\epsilon)$. It is interesting that Brillouin's readout result mirrors the same error–dissipation scaling as found in our analysis, despite the fact that detailed fluctuation theorems, the foundations of our analysis, were still decades away. Brillouin's readout penalty is distinct from our result, but it hints at a more general error–dissipation tradeoff (under particular physical constraints) that could unite the results.

Brillouin's readout penalty also appeared in disguise in Ref. [50], which again, superficially, appears to propose a similar error–dissipation tradeoff to what we have identified. As in Ref. [51], Ref. [50] defines a 'logic operation' specifically as bit inversion, $0 \mapsto 1$ and $1 \mapsto 0$—a strictly reciprocal computation. Reference [50] thus defines the "minimal work per logic operation" as the minimal work for bit inversion. Despite similarities in appearance, it is clearly different from our result since it predicts divergent dissipation for reliable implementation of a strictly-reciprocal computation. In this setting, if the voltage drifts from one input to the other (a readout error), then the wrong input will be inverted. While some computational implementations may require further inefficiency to avoid this source of error, it is distinct and perhaps additive to our nonreciprocity dissipation.

A discussion of reliable computing would be incomplete without mention of von Neumann's well-known analysis of error-correction [52]. Indeed, it might seem plausible that less reliable components can be harnessed to avoid the error–dissipation tradeoff while nevertheless achieving the same net reliability through error correction. However, applying the analysis above to account for the dissipation from all components one finds that error correction via redundancy leads to a *worse* error–dissipation tradeoff. We report on this elsewhere.

Thermodynamically inspired studies of classical reversible computing, e.g., by Bennett, Fredkin, and Tofolli [1, 40],

inspired much early work on quantum computation—e.g., in Refs. [53–55]—where logical reversibility of the unitary transformations is manifest. However, nonreciprocity brings dissipation back to the table, even for reversible logic. We hope it is now clear there is a necessary research program to investigate nonreciprocity dissipation in quantum computing.

Recently, several complementary results suggest that restrictions on control lead to additional work beyond the Landauer bound and so to dissipation. For instance, if constrained to operate in finite time $\tau$, the dissipation using *optimal* protocols scales as the inverse $\tau^{-1}$ of the protocol's duration. The three-way tradeoff between accuracy, speed, and energy efficiency has been explored recently in several settings [11, 45, 56–58]. These suggest that if performed sufficiently slowly, an information processing protocol could achieve the Landauer bound, with zero excess dissipation. In contrast, we found that for time-symmetric protocols, the dissipation scales as $\ln(\epsilon^{-1})$. Thus, one cannot achieve the Landauer bound, no matter how slowly the protocol is executed. In this, our result is more akin to other recently explored control restrictions, such as modularity [5, 33]. This too prevents efficient extraction of nonequilibrium free energy, no matter how slowly the protocol is executed.

## VI. DISCUSSION: TIME ASYMMETRY?

Another insight from the preceding is exploring time-*asymmetric* control when designing energetically efficient computation. However, is time-asymmetry free? Often in thermodynamic control, one posits an external signal that can be produced in any sequence, reversible or not, at no cost beyond the work-energy imparted to the driven system. Inexpensive time asymmetry can be generated, for example, from inertial degrees of freedom [59].

However, in biological systems and in many engineered nanoscale systems, inertial degrees of freedom are not available. Noninertial batteries reliably transforming noninertial systems will drain according to our time-symmetric-control bounds on dissipation.

Time-symmetric control might also seem to apply, rather broadly, if one removes the boundaries separating a driven system from its driver. Including the driver as part of an enlarged system gives the appearance of trivially time-symmetric driving of the composite system. Do our time-symmetric-control bounds on dissipation then apply to any reliable computations that occur within this enlarged system? Generically, our bounds will *not* apply to such enlarged systems because the enlarged microstate distributions $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_\tau$ will no longer be metastable.

This suggests that time-asymmetry and instability can both lead to energetic advantages. Future efforts explore

*which* features of time-asymmetry matter for efficiency. This helps identify those qualitative features of time-symmetric control that must be overcome for improved efficiency.

For time-symmetric protocols, we now ask: How does $\epsilon$ scale with $\tau$ or with the modularity of control? Generically, we do not expect a one-to-one relationship between the error and duration. Rather, generally, there will be a high-dimensional space of tradeoff scalings that includes speed, error, modularity of control, robustness of information storage, energetic efficiency, and many other computationally-relevant design restrictions.

## VII. CONCLUSION

We discovered that time symmetry and metastability together imply a general 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 less than $\epsilon$—the minimum work diverges according to:

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

Of all of this work, only the meager Landauer cost $\Delta\,\mathrm{H}(\mathcal{M}_t)$, which saturates to some finite value as $\epsilon \to 0$, can be recovered in principle. Thus, irretrievable dissipation scales as $\ln(\epsilon^{-1})$. The reciprocity coefficient $\left\langle \left[\!\left[ \mathcal{C}(\mathcal{C}(\mathcal{M}_0)^\dagger)^\dagger \neq \mathcal{M}_0 \right]\!\right] \right\rangle_{\mathcal{M}_0}$ depends only on the deterministic computation to be approximated. This, in turn, identifies sources of energy inefficiency in current implementations of reliable computation. It also implies that time-asymmetric control can allow for more efficient computation—but only when time-asymmetric signals are a free resource.

Restricting to time-symmetric driving may seem unusual. However, with the time-symmetric clock signal that drives modern microprocessors, it is the norm rather than the

exception in the realm of informational controls. Sections IV B and IV C discussed the naturalness of time-symmetry in digital computers and in genetic processing, respectively. To get there, though, required developing several prerequisite thermodynamic results.

After developing the general thermodynamics, we analyzed the limit of highly-reliable computing. As important examples, we then specialized to dissipation in erasure, logic gates, and reliable biological transformations.

Overall, the results elevate time-asymmetric control to a *design principle* for efficient thermodynamic computing. This then must be added to the growing list of recently-discovered principles, including tracking system modularity [5, 33] and logic gates whose protocols adapt to their input [5]. Further progress will turn on how these lessons are incorporated as constraints in the principled design and search for near-optimal finite-time protocols. An opportunity presents itself to adapt these lessons to developing fast hyper-efficient computers in the not-so-distant future.

[1] C. H. Bennett. Thermodynamics of computation—A review. *Intl. J. Theo. Phys.*, 21:905, 1982. 1, 9, 12, 14, 16

[2] O. J. E. Maroney. Generalizing Landauer's principle. *Phys. Rev. E*, 79:031105, Mar 2009. 2, 9, 12

[3] T. Sagawa. Thermodynamic and logical reversibilities revisited. *J. Stat. Mech. Th. Exp.*, 2014(3):P03025, 2014.

[4] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa. Thermodynamics of information. *Nature Physics*, 11(2):131–139, February 2015. 4, 8, 9

[5] 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. 1, 2, 8, 12, 16, 17, 7

[6] P. Salamon and R. S. Berry. Thermodynamic length and dissipated availability. *Phys. Rev. Lett.*, 51:1127–1130, Sep 1983. 1

[7] D. A. Sivak and G. E. Crooks. Thermodynamic metrics and optimal paths. *Phys. Rev. Lett.*, 108:190602, May 2012.

[8] P. R. Zulkowski and M. R. DeWeese. Optimal finite-time erasure of a classical bit. *Phys. Rev. E*, 89(5):052140, 2014.

[9] M. VS. Bonança and S. Deffner. Optimal driving of isothermal processes close to equilibrium. *J. Chem. Phys.*, 140(24):244119, 2014.

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

[11] 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, 16

[12] A. P. Chandrakasan, S. Sheng, and R. W. Brodersen. Low-power CMOS digital design. *IEEE J. Solid-State Circ.*, 27(4):473–484, April 1992. 2

[13] A. Iyer and D. Marculescu. Power efficiency of voltage scaling in multiple clock, multiple voltage cores. In *Proc. 2002 IEEE/ACM Intl. Cong. Computer-aided Design*, pages 379–386. ACM, 2002. 2

[14] J. S. Seldenthuis, F. Prins, J. M. Thijssen, and H. S. J. van der Zant. An all-electric single-molecule motor. *ACS nano*, 4(11):6681–6686, 2010. 2

[15] R. D. Astumian. How molecular motors work–insights from the molecular machinist's toolbox: The Nobel Prize in Chemistry 2016. *Chemical Sci.*, 8(2):840–845, 2017. 2, 14

[16] A. I. Brown and D. A. Sivak. Toward the design principles of molecular machines. *arXiv:1701.04868.* 2

[17] L. Zhang, V. Marcos, and D. A. Leigh. Molecular machines with bio-inspired mechanisms. *Proc. Natl. Acad. Sci.*, 115(38):9397–9404, 2018. 2

[18] E. H. Feng and G. E. Crooks. Length of time's arrow. *Phys. Rev. Lett.*, 101:090602, Aug 2008. 2, 14

[19] R. Landauer. Irreversibility and heat generation in the computing process. *IBM J. Res. Develop.*, 5(3):183–191, 1961. 2, 12

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

[21] C. Jarzynski. Hamiltonian derivation of a detailed fluctuation theorem. *J. Stat. Physics*, 98(1-2):77–102, 2000. 3

[22] U. Seifert. Entropy production along a stochastic trajectory and an integral fluctuation theorem. *Phys. Rev. Lett.*, 95:040602, Jul 2005. 3

[23] M. Esposito and C. Van den Broeck. Second law and Landauer principle far from equilibrium. *Eur. Phys. Lett.*, 95(4), 2011. 3

[24] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley-Interscience, New York, second edition, 2006. 3, 4, 13, 1

[25] M. Esposito. Stochastic thermodynamics under coarse graining. *Phys. Rev. E*, 85(4):041125, 2012. 4, 6

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

[27] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck. Dissipation: The phase-space perspective. *Phys. Rev. Lett.*, 98:080602, 2007. 4

[28] 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. 4

[29] E. Roldan and J. M. R. Parrondo. Estimating dissipation from single stationary trajectories. *Phys. Rev. Lett.*, 105(150607), 2010.

[30] J. Horowitz and C. Jarzynski. Illustrative example of the relationship between dissipation and relative entropy. *Phys. Rev. E*, 79(021106), 2009. 4

[31] E. Roldan and J. M. R. Parrondo. *Estimating dissipation from single stationary trajectories.* PhD thesis, Universidad Complutense de Madrid, 28040 Madrid, Spain, September 2010. 4

[32] Y. Jun, M. Gavrilov, and J. Bechhoefer. High-precision test of Landauer's principle. *Phys. Rev. Lett.*, 113:190601, 2014. 5

[33] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Thermodynamics of modularity: Structural costs beyond the landauer bound. *Phys. Rev. X*, 8(031036), August 2018. 8, 16, 17

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

[35] J. L. England. Statistical physics of self-replication. *J. Chem. Physics*, 139(12):121923, 2013. 8, 4

[36] G. Wimsatt, O.-P. Saira, A. B. Boyd, M. H. Matheny, S. Han, M. L. Roukes, and J. P. Crutchfield. Harnessing fluctuations in thermodynamic computing via time-reversal symmetries. *arXiv:1906.11973.* 9

[37] G. W. Wimsatt, A. B. Boyd, P. M. Riechers, and J. P. Crutchfield. Balancing error and dissipation when erasing information. *arXiv:2002.XXXXX.* 12, 13

[38] P. Shivakumar, M. Kistler, S. W. Keckler, D. Burger, and L. Alvisi. Modeling the effect of technology trends on the soft error rate of combinational logic. In *Dependable Systems and Networks, 2002. DSN 2002. Proceedings. International Conference on*, pages 389–398. IEEE, 2002. 14

[39] E. Fredkin and T. Toffoli. Conservative logic. *Intl. J. Theo. Phys.*, 21(3/4):219–253, 1982. 14

[40] E. Fredkin and T. Toffoli. Conservative logic. In *Collision-based computing*, pages 47–81. Springer, 2002. 14, 16

[41] P. M. Riechers and J. P. Crutchfield. Fluctuations when driving between nonequilibrium steady states. *J. Stat. Physics*, 168(4):873–918, 2017. 14

[42] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England. Dissipation bounds all steady-state current fluctuations. *Phys. Rev. Lett.*, 116:120601, Mar 2016. 15

[43] J. J. Hopfield. Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci.*, 71(10):4135–4139, 1974. 15

[44] Charles H. Bennett. Dissipation-error tradeoff in proofreading. *Biosystems*, 11(2):85 – 91, 1979.

[45] A. Murugan, D. A. Huse, and S. Leibler. Speed, dissipation, and error in kinetic proofreading. *Proc. Natl. Acad. Sci.*, 109(30):12034–12039, 2012. 16

[46] P. Sartori and S. Pigolotti. Thermodynamics of error correction. *Phys. Rev. X*, 5:041039, Dec 2015. 15

[47] J. M. Horowitz, K. Zhou, and J. L. England. Minimum energetic cost to maintain a target nonequilibrium state. *Phys. Rev. E*, 95:042102, Apr 2017. 15

[48] T. E. Ouldridge, C. C. Govern, and P. R. ten Wolde. Thermodynamics of computational copying in biochemical systems. *Phys. Rev. X*, 7(2):021004, 2017. 15

[49] L. Brillouin. *Science and Information Theory.* Academic Press, New York, 1956. 16

[50] K.-L. Stein. Noise-induced error rate as limiting factor for energy per operation in digital ICs. *IEEE J. Solid State Circuits*, SC-12(5):527–530, 1977. 16

[51] R. Muller, H.J. Pfleiderer, and K.U. Stein. Energy per logic operation in integrated circuits: Definition and determination. *IEEE J. Solid-State Circ.*, 11(5):657–661, 1976. 16

[52] J. von Neumann. Probabilistic logics and the synthesis of reliable organisms from unreliable components. In C. E. Shannon and J. McCarthy, editors, *Automata Studies*, number 34 in Annals of Mathematical Studies, pages 329–378. Princeton University Press, Princeton, New Jersey, 1956. 16

[53] Paul A. Benioff. Quantum mechanical hamiltonian models of discrete processes that erase their own histories: Application to turing machines. *Intl. J. Theo. Phys.*, 21(3-4):177–201, 1982. 16

[54] W. H. Zurek. Reversibility and stability of information processing systems. *Physical Review Letters*, 53(4):391, 1984.

[55] A. Peres. Reversible logic and quantum computers. *Phys. Rev. A*, 32:3266–3276, Dec 1985. 16

[56] 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. 16

[57] P. R. Zulkowski and M. R. DeWeese. Optimal finite-time erasure of a classical bit. *Phys. Rev. E*, 89:052140, 2014.

[58] S. Lahiri, J. Sohl-Dickstein, and S. Ganguli. A universal tradeoff between power, precision and speed in physical communication. *arXiv:1603.07758*. 16

[59] S. Deffner and C. Jarzynski. Information processing and the second law of thermodynamics: An inclusive, Hamiltonian approach. *Phys. Rev. X*, 3:041003, 2013. 16

[60] J. A. Owen, A. Kolchinsky, and D. H. Wolpert. Number of hidden states needed to physically implement a given conditional distribution. *New J. Physics*, 21:013022, 2018. 6

Supplementary Materials

*Balancing Error and Dissipation in Computing*

Paul M. Riechers, Alexander B. Boyd, Gregory W. Wimsatt, and James P. Crutchfield

The Supplementary Material derives a general bound on dissipated work in thermodynamic computing, gives a transition-specific fluctuation theorem, details the four cases of time-symmetric memories, and analyzes the minimal dissipation in the time-symmetric universal NAND gate.

**Appendix A: Dissipated-work bound for computation**

By the information-processing inequality [24], the relative entropy between two distributions is nonincreasing under a general channel applied to both distributions. This has important implications for thermodynamic entropy production.

For our main results, we consider the case where we observe memory states $\boldsymbol{\mathcal{M}}$, where $\boldsymbol{\mathcal{M}}$ is a partitioning of the microstates $\boldsymbol{\mathcal{S}}$, at the initial and final times. The random variable $\mathcal{M}_t$ for the memory state at time $t$ is completely determined by the random variable for the microstate $\mathcal{S}_t$ via the memory coarse-graining function $f_{\mathcal{M}} \colon \boldsymbol{\mathcal{S}} \to \boldsymbol{\mathcal{M}}$. In the following, we also invoke an even coarser graining of microstate trajectories which takes microstate trajectories to the pair of starting and ending memory states: $f_{\mathcal{M}_0, \mathcal{M}_\tau}(s_{0:\tau}) = \big(f_{\mathcal{M}}(s_0), f_{\mathcal{M}}(s_\tau)\big)$.

We begin by deriving a bound on dissipation in terms of a general partitioning $\boldsymbol{\mathcal{M}}$.

Note that the two probability distributions $\rho(s_{0:\tau}) \equiv \Pr_{x_{0:\tau}}\big(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{S}_0 \sim \boldsymbol{\mu}_0\big)$ and $\rho^R(s_{0:\tau}) \equiv \Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})}\big(\mathcal{S}_{0:\tau} = \boldsymbol{\mathcal{A}}(s_{0:\tau}) | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger\big)$ each have the same support $s_{0:\tau} \in \boldsymbol{\mathcal{S}}^{[0,\tau]}$. From Eq. (3), the expected dissipated work is: $\langle W_{\text{diss}} \rangle = k_{\text{B}} T \, \text{D}_{\text{KL}} \big[\rho \, || \, \rho^R\big]$.

Now, consider the coarse-grained marginal distributions: $\boldsymbol{P}(m, m') = \sum_{s_{0:\tau}} \delta_{s_0 \in m} \delta_{s_\tau \in m'} \, \rho(s_{0:\tau})$ and $\boldsymbol{Q}(m, m') = \sum_{s_{0:\tau}} \delta_{s_0 \in m} \delta_{s_\tau \in m'} \, \rho^R(s_{0:\tau})$.

Since the same coarse graining kernel from $\boldsymbol{\mathcal{S}}^{[0,\tau]}$ to $\boldsymbol{\mathcal{M}}^2$ is applied to each of the distributions $\rho$ and $\rho^R$, the data-processing inequality guarantees that: $\text{D}_{\text{KL}}\big[\boldsymbol{P} \, || \, \boldsymbol{Q}\big] \le \text{D}_{\text{KL}}\big[\rho \, || \, \rho^R\big]$.

Moreover, note that:

$$\boldsymbol{P}(m, m') = \sum_{s_{0:\tau}} \delta_{s_0 \in m} \delta_{s_\tau \in m'} \Pr_{x_{0:\tau}} \big(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{S}_0 \sim \boldsymbol{\mu}_0\big)$$
$$= \Pr_{x_{0:\tau}} \big(\mathcal{M}_0 = m, \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0\big) \ ,$$

and:

$$\boldsymbol{Q}(m, m') = \sum_{s_{0:\tau}} \delta_{s_0 \in m} \delta_{s_\tau \in m'} \Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \big(\mathcal{S}_{0:\tau} = \boldsymbol{\mathcal{A}}(s_{0:\tau}) | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger\big)$$
$$= \Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \big(\mathcal{S}_0 \in m'^\dagger, \mathcal{S}_\tau \in m^\dagger | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger\big) \ .$$

Hence, this shows in effect the dissipated work is always lower-bounded by a function of the net transition probabilities between memory states:

$$\beta \langle W_{\text{diss}} \rangle = \text{D}_{\text{KL}} \left[ \Pr_{x_{0:\tau}} \big(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{S}_0 \sim \boldsymbol{\mu}_0\big) \ \middle|\middle| \ \Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \big(\mathcal{S}_{0:\tau} = \boldsymbol{\mathcal{A}}(s_{0:\tau}) | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger\big) \right]$$
$$\ge \text{D}_{\text{KL}} \left[ \Pr_{x_{0:\tau}} \big(\mathcal{M}_0 = m, \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0\big) \ \middle|\middle| \ \Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \big(\mathcal{S}_0 \in m'^\dagger, \mathcal{S}_\tau \in m^\dagger | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger\big) \right] \ . \tag{S1}$$

Equation (S1) is valid even when $m^\dagger \notin \boldsymbol{\mathcal{M}}$.

When $m^\dagger \in \boldsymbol{\mathcal{M}}$ for all $m \in \boldsymbol{\mathcal{M}}$, Eq. (S1) reduces to:

$$\beta \langle W_{\text{diss}} \rangle \ge \text{D}_{\text{KL}} \left[ \Pr_{x_{0:\tau}} \big(\mathcal{M}_0 = m, \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0\big) \ \middle|\middle| \ \Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \big(\mathcal{M}_0 = m'^\dagger, \mathcal{M}_\tau = m^\dagger | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger\big) \right] \ . \tag{S2}$$

Equation (S2) will be broadly applicable to computations performed on positional, configurational, or typical magnetic memory systems. Indeed, there are many standard types of memory states which satisfy $m^\dagger \in \boldsymbol{\mathcal{M}}$, so that Eq. (S2) is applicable.

For example, positionally partitioned states are often used to store information in systems with potential energy minima at different spatial locations. (As implemented in, for example, Maxwell demon experiments via laser ion

traps, also called 'optical tweezers' when the location of these traps is dynamically controlled.) These positional memories have a clear definition of time reversal in which $m^\dagger = m$. This follows since if $s = (\vec{q}, \vec{\wp})$ is in $m$, then so is $s^\dagger = (\vec{q}, -\vec{\wp})$. Another form of memory state utilizes magnetic spins, for which the time reversal flips all microstate spins—spin-up maps to spin-down upon time-reversal, and vice versa. In the case of bistable magnetic memory elements, the time-reversal of any memory is also a valid memory $m^\dagger \in \mathcal{M}$, although $m^\dagger \neq m$. Thus, for both of these forms of memory, the reversal of a memory state is an element of the partitioning, so that Eq. (S2) is applicable. Further decomposing Eq. (S2), we find:

$$\beta \langle W_{\text{diss}} \rangle \geq \mathrm{D}_{\mathrm{KL}} \left[ \Pr_{x_{0:\tau}} \left( \mathcal{M}_0 = m, \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0 \right) \,\bigg\|\, \Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \left( \mathcal{M}_0 = m'^\dagger, \mathcal{M}_\tau = m^\dagger | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger \right) \right]$$

$$= \sum_{m,m' \in \mathcal{M}} \Pr_{x_{0:\tau}} \left( \mathcal{M}_0 = m, \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0 \right) \ln \left( \frac{\Pr_{x_{0:\tau}} \left( \mathcal{M}_0 = m, \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0 \right)}{\Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \left( \mathcal{M}_0 = m'^\dagger, \mathcal{M}_\tau = m^\dagger | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger \right)} \right)$$

$$= \sum_{m,m' \in \mathcal{M}} \Pr_{x_{0:\tau}} \left( \mathcal{M}_0 = m, \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0 \right)$$

$$\times \ln \left( \frac{\Pr_{x_{0:\tau}} \left( \mathcal{M}_0 = m | \mathcal{S}_0 \sim \boldsymbol{\mu}_0 \right) \Pr_{x_{0:\tau}} \left( \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0, \mathcal{M}_0 = m \right)}{\Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \left( \mathcal{M}_0 = m'^\dagger | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger \right) \Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \left( \mathcal{M}_\tau = m^\dagger | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger, \mathcal{M}_0 = m'^\dagger \right)} \right)$$

$$= \sum_{m,m' \in \mathcal{M}} \Pr_{x_{0:\tau}} \left( \mathcal{M}_0 = m, \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0 \right) \ln \left( \Pr_{x_{0:\tau}} \left( \mathcal{M}_0 = m | \mathcal{S}_0 \sim \boldsymbol{\mu}_0 \right) \right)$$

$$- \sum_{m,m' \in \mathcal{M}} \Pr_{x_{0:\tau}} \left( \mathcal{M}_0 = m, \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0 \right) \ln \left( \Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \left( \mathcal{M}_0 = m'^\dagger | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger \right) \right)$$

$$+ \sum_{m,m' \in \mathcal{M}} \Pr_{x_{0:\tau}} \left( \mathcal{M}_0 = m, \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0 \right) \ln \left( \frac{\Pr_{x_{0:\tau}} \left( \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0, \mathcal{M}_0 = m \right)}{\Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \left( \mathcal{M}_\tau = m^\dagger | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger, \mathcal{M}_0 = m'^\dagger \right)} \right)$$

$$= \sum_{m \in \mathcal{M}} \Pr_{x_{0:\tau}} \left( \mathcal{M}_0 = m | \mathcal{S}_0 \sim \boldsymbol{\mu}_0 \right) \ln \left( \Pr_{x_{0:\tau}} \left( \mathcal{M}_0 = m | \mathcal{S}_0 \sim \boldsymbol{\mu}_0 \right) \right)$$

$$- \sum_{m' \in \mathcal{M}} \Pr_{x_{0:\tau}} \left( \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0 \right) \ln \left( \Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \left( \mathcal{M}_0 = m'^\dagger | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger \right) \right)$$

$$+ \sum_{m,m' \in \mathcal{M}} \Pr_{x_{0:\tau}} \left( \mathcal{M}_0 = m | \mathcal{S}_0 \sim \boldsymbol{\mu}_0 \right) \Pr_{x_{0:\tau}} \left( \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0, \mathcal{M}_0 = m \right) \ln \left( \frac{\Pr_{x_{0:\tau}} \left( \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0, \mathcal{M}_0 = m \right)}{\Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \left( \mathcal{M}_\tau = m^\dagger | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^\dagger, \mathcal{M}_0 = m'^\dagger \right)} \right)$$

$$= \sum_{m \in \mathcal{M}} \boldsymbol{\mu}_0(m) \ln \left( \boldsymbol{\mu}_0(m) \right) - \sum_{m' \in \mathcal{M}} \boldsymbol{\mu}_\tau(m') \ln \left( \boldsymbol{\mu}_\tau^\dagger(m'^\dagger) \right)$$

$$+ \sum_{m,m' \in \mathcal{M}} \boldsymbol{\mu}_0(m) \Pr_{x_{0:\tau}} \left( \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0^{(m)} \right) \ln \left( \frac{\Pr_{x_{0:\tau}} \left( \mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0^{(m)} \right)}{\Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \left( \mathcal{M}_\tau = m^\dagger | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^{\dagger(m'^\dagger)} \right)} \right)$$

$$= - \mathrm{H}(\mathcal{M}_0) + \mathrm{H}(\mathcal{M}_\tau) + \sum_{m,m' \in \mathcal{M}} \boldsymbol{\mu}_0(m) d(m, m')$$

$$= \Delta \mathrm{H}(\mathcal{M}_t) + \sum_{m,m' \in \mathcal{M}} \boldsymbol{\mu}_0(m) d(m, m') \ . \tag{S3}$$

In the third-to-last equation above we used the fact that $\boldsymbol{\mu}_\tau^\dagger(m^\dagger) = \boldsymbol{\mu}_\tau(m)$ and we defined:

$$d(m, m') \equiv \Pr_{x_{0:\tau}} \left( \mathcal{M}_\tau = m' \big| \mathcal{S}_0 \sim \boldsymbol{\mu}_0^{(m)} \right) \ln \left( \frac{\Pr_{x_{0:\tau}} \left( \mathcal{M}_\tau = m' \big| \mathcal{S}_0 \sim \boldsymbol{\mu}_0^{(m)} \right)}{\Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \left( \mathcal{M}_\tau = m^\dagger \big| \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^{\dagger(m'^\dagger)} \right)} \right) \ .$$

A more general derivation (*not* assuming that $m^\dagger \in \mathcal{M}$), fully leveraging Eq. (S1), yields Eq. (S3) with:

$$d(m, m') \equiv \Pr_{x_{0:\tau}} \left( \mathcal{M}_\tau = m' \big| \mathcal{S}_0 \sim \boldsymbol{\mu}_0^{(m)} \right) \ln \left( \frac{\Pr_{x_{0:\tau}} \left( \mathcal{M}_\tau = m' \big| \mathcal{S}_0 \sim \boldsymbol{\mu}_0^{(m)} \right)}{\Pr_{\boldsymbol{\mathcal{A}}(x_{0:\tau})} \left( \mathcal{S}_\tau \in m^\dagger \big| \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^{\dagger(m'^\dagger)} \right)} \right) \ .$$

This generalization is useful, for example, when analyzing tristable magnetic memory elements.

In the $m = m^\dagger$ case, where memories are stored via time-symmetric variables, these expressions both simplify to:

$$d(m, m') = \Pr_{x_{0:\tau}}\left(\mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0^{(m)}\right) \ln \left( \frac{\Pr_{x_{0:\tau}}\left(\mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\mu}_0^{(m)}\right)}{\Pr_{\boldsymbol{A}(x_{0:\tau})}\left(\mathcal{M}_\tau = m | \mathcal{S}_0 \sim \boldsymbol{\mu}_\tau^{\dagger(m')}\right)} \right) \ .$$

This compares (i) the probability of transitioning from memory state $m$ to $m'$ to (ii) the probability of returning back to $m$ upon reversal of time-odd variables (like momentum) and subsequent reversal of the control protocol. Above, $\boldsymbol{\mu}_t^{(m)} \equiv \delta_{\mathcal{S}_t \in m} \boldsymbol{\mu}_t / \sum_{s' \in m} \boldsymbol{\mu}_t(s')$ is the renormalized restriction of the microstate distribution $\boldsymbol{\mu}_t$ to the set of microstates represented by the memory $m$. Note that (despite the fact that $\boldsymbol{\mu}_\tau^\dagger(m^\dagger) = \boldsymbol{\mu}_\tau(m)$), generically $\boldsymbol{\mu}_\tau^{\dagger(m^\dagger)} \neq \boldsymbol{\mu}_\tau^{(m)}$. However, conveniently, metastable memory systems have: $\boldsymbol{\mu}_\tau^{\dagger(m)} \approx \boldsymbol{\pi}_{x_\tau}^{\dagger(m)} = \boldsymbol{\pi}_{x_\tau^\dagger}^{(m)}$.

## Appendix B: Transition-specific fluctuation theorem

Here, we verify the transition-specific version of Eq. (18). We start with the detailed fluctuation theorem for heat:

$$e^{\beta Q(s_{0:\tau}, x_{0:\tau})} = \frac{\Pr_{x_{0:\tau}}\left(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{S}_0 = s_0\right)}{\Pr_{\boldsymbol{A}(x_{0:\tau})}\left(\mathcal{S}_{0:\tau} = \boldsymbol{A}(s_{0:\tau}) | \mathcal{S}_0 = s_\tau^\dagger\right)} \ .$$

We then apply the First Law of thermodynamics (energy conservation)—$E_{x_\tau}(s_\tau) - E_{x_0}(s_0) = W - Q$—together with a useful identity: $E_x(s) = -k_B T \ln\left(\boldsymbol{\pi}_x^{(m)}(s)\right) + F_x^{(m)}$, if $s \in m$. This allows us to write:

$$W - \Delta F_x^{(m)} = Q - k_B T \Delta \ln\left(\boldsymbol{\pi}_x^{(m)}(s)\right) \ .$$

Then we calculate:

$$\left\langle e^{-\beta(W - \Delta F_x^{(\mathcal{M})})} \right\rangle_{\Pr_{x_{0:\tau}}(\mathcal{S}_{0:\tau} | \mathcal{M}_\tau = m', \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_0}^{(m)})}$$

$$= \left\langle e^{-\beta\left(Q - k_B T \Delta \ln\left(\boldsymbol{\pi}_x^{(\mathcal{M})}(\mathcal{S})\right)\right)} \right\rangle_{\Pr_{x_{0:\tau}}(\mathcal{S}_{0:\tau} | \mathcal{M}_\tau = m', \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_0}^{(m)})}$$

$$= \sum_{s_{0:\tau} \in \boldsymbol{S}^{[0:\tau]}} \Pr_{x_{0:\tau}}\left(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{M}_\tau = m', \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_0}^{(m)}\right) \frac{\boldsymbol{\pi}_{x_\tau}^{(m')}(s_\tau)}{\boldsymbol{\pi}_{x_0}^{(m)}(s_0)} \frac{\Pr_{\boldsymbol{A}(x_{0:\tau})}\left(\mathcal{S}_{0:\tau} = \boldsymbol{A}(s_{0:\tau}) | \mathcal{S}_0 = s_\tau^\dagger\right)}{\Pr_{x_{0:\tau}}\left(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{S}_0 = s_0\right)}$$

$$= \sum_{s_{0:\tau} \in m\boldsymbol{S}^{(0:\tau)}m'} \Pr_{x_{0:\tau}}\left(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{M}_\tau = m', \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_0}^{(m)}\right) \frac{\boldsymbol{\pi}_{x_\tau}^{(m')}(s_\tau)}{\boldsymbol{\pi}_{x_0}^{(m)}(s_0)} \frac{\Pr_{\boldsymbol{A}(x_{0:\tau})}\left(\mathcal{S}_{0:\tau} = \boldsymbol{A}(s_{0:\tau}) | \mathcal{S}_0 = s_\tau^\dagger\right)}{\Pr_{x_{0:\tau}}\left(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{S}_0 = s_0\right)}$$

$$= \frac{1}{\Pr_{x_{0:\tau}}(\mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_0}^{(m)})} \sum_{s_{0:\tau} \in m\boldsymbol{S}^{(0:\tau)}m'} \Pr_{x_{0:\tau}}\left(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_0}^{(m)}\right) \frac{\boldsymbol{\pi}_{x_\tau}^{(m')}(s_\tau)}{\boldsymbol{\pi}_{x_0}^{(m)}(s_0)} \frac{\Pr_{\boldsymbol{A}(x_{0:\tau})}\left(\mathcal{S}_{0:\tau} = \boldsymbol{A}(s_{0:\tau}) | \mathcal{S}_0 = s_\tau^\dagger\right)}{\Pr_{x_{0:\tau}}\left(\mathcal{S}_{0:\tau} = s_{0:\tau} | \mathcal{S}_0 = s_0\right)}$$

$$= \frac{1}{\Pr_{x_{0:\tau}}(\mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_0}^{(m)})} \sum_{s_{0:\tau} \in m\boldsymbol{S}^{(0:\tau)}m'} \boldsymbol{\pi}_{x_\tau}^{(m')}(s_\tau) \Pr_{\boldsymbol{A}(x_{0:\tau})}\left(\mathcal{S}_{0:\tau} = \boldsymbol{A}(s_{0:\tau}) | \mathcal{S}_0 = s_\tau^\dagger\right)$$

$$= \frac{1}{\Pr_{x_{0:\tau}}(\mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_0}^{(m)})} \sum_{\boldsymbol{A}(s_{0:\tau}) \in m'^\dagger \boldsymbol{S}^{(0:\tau)}m^\dagger} \boldsymbol{\pi}_{x_\tau^\dagger}^{(m'^\dagger)}(s_\tau^\dagger) \Pr_{\boldsymbol{A}(x_{0:\tau})}\left(\mathcal{S}_{0:\tau} = \boldsymbol{A}(s_{0:\tau}) | \mathcal{S}_0 = s_\tau^\dagger\right)$$

$$= \frac{1}{\Pr_{x_{0:\tau}}(\mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_0}^{(m)})} \sum_{\boldsymbol{A}(s_{0:\tau}) \in m'^\dagger \boldsymbol{S}^{(0:\tau)}m^\dagger} \Pr_{\boldsymbol{A}(x_{0:\tau})}\left(\mathcal{S}_{0:\tau} = \boldsymbol{A}(s_{0:\tau}) | \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_\tau^\dagger}^{(m'^\dagger)}\right)$$

$$= \frac{1}{\Pr_{x_{0:\tau}}(\mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_0}^{(m)})} \sum_{s \in m^\dagger} \Pr_{\boldsymbol{A}(x_{0:\tau})}\left(\mathcal{S}_\tau = s | \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_\tau^\dagger}^{(m'^\dagger)}\right)$$

$$= \frac{\Pr_{\boldsymbol{A}(x_{0:\tau})}\left(\mathcal{M}_\tau = m^\dagger | \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_\tau^\dagger}^{(m'^\dagger)}\right)}{\Pr_{x_{0:\tau}}(\mathcal{M}_\tau = m' | \mathcal{S}_0 \sim \boldsymbol{\pi}_{x_0}^{(m)})} \ ,$$

where we used the fact that $\boldsymbol{\pi}_{x_\tau}^{(m')}(s_\tau) = \boldsymbol{\pi}_{x_\tau^\dagger}^{(m'^\dagger)}(s_\tau^\dagger)$.

Altogether, this leads to a useful transition-specific fluctuation theorem:

$$\left\langle e^{-\beta W}\right\rangle_{\mathrm{Pr}_{x_{0:\tau}}\left(\mathcal{S}_{0:\tau}|\mathcal{M}_\tau=m',\mathcal{S}_0\sim\boldsymbol{\pi}_{x_0}^{(m)}\right)} = e^{-\beta\Delta F_x^{(\mathcal{M})}} \frac{\mathrm{Pr}_{\boldsymbol{\mathcal{A}}(x_{0:\tau})}\left(\mathcal{M}_\tau=m^\dagger|\mathcal{S}_0\sim\boldsymbol{\pi}_{x_\tau^\dagger}^{(m'^\dagger)}\right)}{\mathrm{Pr}_{x_{0:\tau}}\left(\mathcal{M}_\tau=m'|\mathcal{S}_0\sim\boldsymbol{\pi}_{x_0}^{(m)}\right)} \ .$$

By Jensen's inequality, this yields:

$$\left\langle W\right\rangle_{\mathrm{Pr}_{x_{0:\tau}}\left(\mathcal{S}_{0:\tau}|\mathcal{M}_\tau=m',\mathcal{S}_0\sim\boldsymbol{\pi}_{x_0}^{(m)}\right)} \geq \Delta F_x^{(\mathcal{M})} + k_\mathrm{B}T\ln\frac{\mathrm{Pr}_{x_{0:\tau}}\left(\mathcal{M}_\tau=m'|\mathcal{S}_0\sim\boldsymbol{\pi}_{x_0}^{(m)}\right)}{\mathrm{Pr}_{\boldsymbol{\mathcal{A}}(x_{0:\tau})}\left(\mathcal{M}_\tau=m^\dagger|\mathcal{S}_0\sim\boldsymbol{\pi}_{x_\tau^\dagger}^{(m'^\dagger)}\right)} \ .$$

Assuming time-symmetric control, this then implies that the minimal average work required of a memory transition is:

$$W_{\min}^{t\text{-sym}}(m\to m') = F_{x_0}^{(m')} - F_{x_0}^{(m)} + k_\mathrm{B}T\Psi(m\to m') \ .$$

These results exhibit similarity with those of Ref. [35], although the appearance of the local-equilibrium free energies is a new feature here. This has important implications for the work required of metastable memory transitions.

## Appendix C: Four cases

To evaluate Eq. (14) there are four cases to consider. Assuming time-symmetric memories (such that $m^\dagger = m$), the four cases depend on whether $\mathcal{C}(m) = m'$ or $\neq m'$, and on whether $\mathcal{C}(m') = m$ or $\neq m$. Whatever implementation is used for the computation, it will result in some *actual* probability of error for each of the intended transitions: $\mathrm{Pr}_{x_{0:\tau}}\left(\mathcal{M}_\tau = \mathcal{C}(m)|\mathcal{M}_0 = m\right) = 1 - \epsilon_m$. The reliability design constraint is that $\epsilon_m \leq \epsilon$ for all possible initial memories $m$, and that $\epsilon \ll 1$, so that errors are very improbable.

Given some implementation $x_{0:\tau}$, the actual probability of making a particular *accidental* memory transition is $\epsilon_{m\to m'}$ for $m' \neq \mathcal{C}(m)$. Since $\sum_{m'\in\mathcal{M}\setminus\{\mathcal{C}(m)\}} \epsilon_{m\to m'} = \epsilon_m \leq \epsilon$, we must have that $0 \leq \epsilon_{m\to m'} \leq \epsilon_m \leq \epsilon$. Let us use this to evaluate the four possible cases for $d(m, m')$:

1. $\mathcal{C}(m) = m'$; $\mathcal{C}(m') = m$:

$$\begin{aligned}
d^{(1)}(m,m') &= (1-\epsilon_m)\ln\left(\frac{1-\epsilon_m}{1-\epsilon_{m'}}\right) \\
&\geq (1-\epsilon_m)\ln(1-\epsilon_m) \\
&\geq (1-\epsilon)\ln(1-\epsilon) \\
&= (1-\epsilon)\left(-\epsilon - \tfrac{1}{2}\epsilon^2 - \tfrac{1}{3}\epsilon^3 - \dots\right) \\
&= -\epsilon + \tfrac{1}{2}\epsilon^2 + \tfrac{1}{6}\epsilon^3 + \dots \\
&\geq -\epsilon + \tfrac{1}{2}\epsilon^2 \\
&\geq -\epsilon \\
&\approx 0
\end{aligned}$$

Similarly,

$$\begin{aligned}
d^{(1)}(m,m') &= (1-\epsilon_m)\ln\left(\frac{1-\epsilon_m}{1-\epsilon_{m'}}\right) \\
&\leq -\ln(1-\epsilon_{m'}) \\
&\leq -\ln(1-\epsilon) \\
&= \epsilon + \tfrac{1}{2}\epsilon^2 + \tfrac{1}{3}\epsilon^3 + \dots \\
&\approx 0
\end{aligned}$$

So $-\epsilon \leq d^{(1)}(m,m') \leq \epsilon + \tfrac{1}{2}\epsilon^2 + \mathcal{O}(\epsilon^3)$.

2. $\mathcal{C}(m) = m'; \mathcal{C}(m') \neq m$:

$$d^{(2)}(m, m') = (1 - \epsilon_m) \ln \left( \frac{1 - \epsilon_m}{\epsilon_{m' \to m}} \right)$$
$$\geq (1 - \epsilon) \ln \left( \frac{1 - \epsilon}{\epsilon_{m' \to m}} \right)$$
$$\geq (1 - \epsilon) \ln \left( \frac{1 - \epsilon}{\epsilon} \right)$$
$$\approx \ln (1/\epsilon)$$

We also have:

$$d^{(2)}(m, m') = (1 - \epsilon_m) \ln \left( \frac{1 - \epsilon_m}{\epsilon_{m' \to m}} \right)$$
$$\leq \ln(1/\epsilon_{m' \to m})$$

So, $\ln(\epsilon^{-1}) \lesssim d^{(2)}(m, m') \leq \ln(\epsilon_{m' \to m}^{-1})$.

3. $\mathcal{C}(m) \neq m'; \mathcal{C}(m') = m$:

$$d^{(3)}(m, m') = \epsilon_{m \to m'} \ln \left( \frac{\epsilon_{m \to m'}}{1 - \epsilon_{m'}} \right)$$
$$> \epsilon_{m \to m'} \ln \epsilon_{m \to m'}$$
$$\approx 0$$

Also:

$$d^{(3)}(m, m') = \epsilon_{m, m'} \ln \left( \frac{\epsilon_{m, m'}}{1 - \epsilon_{m'}} \right)$$
$$< 0 .$$

So, $\epsilon_{m \to m'} \ln \epsilon_{m \to m'} < d^{(3)}(m, m') < 0$.

4. $\mathcal{C}(m) \neq m'; \mathcal{C}(m') \neq m$:

$$d^{(4)}(m, m') = \epsilon_{m \to m'} \ln \left( \frac{\epsilon_{m \to m'}}{\epsilon_{m' \to m}} \right)$$
$$< \epsilon_{m \to m'} \ln \left( \frac{1}{\epsilon_{m' \to m}} \right)$$
$$\approx 0 ,$$

assuming that the errors $\epsilon_{m \to m'}$ and $\epsilon_{m' \to m}$ are on the same order of magnitude. Also:

$$d^{(4)}(m, m') = \epsilon_{m \to m'} \ln \left( \frac{\epsilon_{m \to m'}}{\epsilon_{m' \to m}} \right)$$
$$> \epsilon_{m \to m'} \ln \epsilon_{m \to m'}$$
$$\approx 0 .$$

So, $\epsilon_{m \to m'} \ln \epsilon_{m \to m'} < d^{(4)}(m, m') < -\epsilon_{m \to m'} \ln \epsilon_{m' \to m}$.

Case 2 is the only one resulting in divergent dissipation with increasing reliability.

However, due to the ratio of error rates in Case 4, its behavior in the low-error limit deserves further analysis.

On the one hand, suppose $\epsilon_{m'\to m} > \epsilon_{m\to m'}$. Let $r \equiv \epsilon_{m\to m'}/\epsilon_{m'\to m} < 1$. Then:

$$d^{(4)}(m, m') = \epsilon_{m\to m'} \ln \frac{\epsilon_{m\to m'}}{\epsilon_{m'\to m}} = \epsilon_{m'\to m} r \ln r .$$

If we define $f(r) = r \ln r$, then $f'(r) = \ln r + 1$. And so, $f'(r^*) = 0$ gives $r^* = e^{-1}$ and $f(r^*) = -e^{-1}$. Since $f(0) = f(1) = 0$, $f$'s local minimum is at $e^{-1}$. So:

$$d^{(4)}(m, m') = \epsilon_{m'\to m} r \ln r \geq -\epsilon_{m'\to m} e^{-1} .$$

However, $\epsilon \geq \epsilon_{m'\to m}$, so $d^{(4)}(m, m') \geq -\epsilon e^{-1}$. And, since $f(r) \leq 0$ for $0 \leq r \leq 1$, we have $0 \geq d^{(4)}(m, m') \geq -\epsilon e^{-1}$. If, on the other hand, $\epsilon_{m\to m'} > \epsilon_{m'\to m}$, then Case 4 will imply some extra dissipation, although it will still be on the order of $\epsilon$ so long as $\epsilon_{m\to m'}$ and $\epsilon_{m'\to m}$ are within several orders of magnitude.

## Appendix D: Minimal dissipation in the time-symmetric universal NAND gate

Let us consider the implications for the common logic gates that serve as the building blocks for practical computers. Recall that the simple NAND gate is sufficient for universal computation. It is therefore worthwhile to consider what dissipation is commonly incurred in such important logic gates—and consider how this dissipation can be avoided.

To address this, consider a physical implementation of the memory components of a NAND gate, where we explicitly consider two memory elements whose memory states are to be used as the input for the NAND gate and another memory element that will store the value of the output. We assume that only the output is over-written during the computation—the input memory states may be kept around for later use.

Note that this is already a particular physical model of the NAND computation—indeed, alternatives exist such as storing the output in the location of one of the former inputs by over-writing one of the inputs. However, we analyze the proposed two-input–one-output model here since it is arguably the most relevant to the typical desired use of a NAND gate. Other ancillary memory elements may be used in the computation as in Ref. [60] but, since they will be returned to their original state by the computation's end, these ancillary memories do not need to result in any explicit additional dissipation dependence—although they can influence the dissipation through their implicit effect on the net errors in operating on the memory-elements of interest—and can be left implicit in the self-consistency of the current analysis.

Each of the three explicitly-considered memory elements is assumed to be bistable; i.e., each has two metastable regions of state space. Let the microstate of each memory element be specified by its position in the interval $(-\pi, \pi]$. Between computations, including at $t = 0$ and $t = \tau$, the metastable regions for each memory element are $0 \equiv (-\pi, 0]$ and $1 \equiv (0, \pi]$ which gives a natural partition for the memory states.

The microstate of the memory system at any time $t$ can be treated as a composite random variable $\mathcal{S}_t = (\mathcal{S}_t^{(\mathrm{in}_1)}, \mathcal{S}_t^{(\mathrm{in}_2)}, \mathcal{S}_t^{(\mathrm{out})})$ with $\mathcal{S}_t^{(\cdot)} \in (-\pi, \pi]$. Similarly, the memory state is the composite random variable $\mathcal{M}_t = (\mathcal{M}_t^{(\mathrm{in}_1)}, \mathcal{M}_t^{(\mathrm{in}_2)}, \mathcal{M}_t^{(\mathrm{out})})$ with $\mathcal{M}_t^{(\cdot)} \in \{0, 1\}$ corresponding to the two metastable regions for each memory element. Thus, the joint state-space $\mathcal{S} = \mathbb{R}^3_{(-\pi,\pi]}$ has eight metastable regions, which we identify as the joint memory system's eight memory states: $\mathcal{M} = \{m_{000}, m_{001}, m_{010}, \ldots m_{111}\}$ where each memory state is labeled according to its corresponding region of state-space: $m_{jk\ell} = \{s \in \mathcal{S} : s^{(\mathrm{in}_1)} \in (-\pi + j\pi, j\pi], s^{(\mathrm{in}_2)} \in (-\pi + k\pi, k\pi], s^{(\mathrm{out})} \in (-\pi + \ell\pi, \ell\pi]\}$. That is, each of the memory states is one of the octants of state space.

$\mathcal{C}_{\mathrm{NAND}}$ is given by the mappings:

$$000 \mapsto 001 \quad 001 \mapsto 001 \quad 010 \mapsto 011 \quad 011 \mapsto 011 \quad 100 \mapsto 101 \quad 101 \mapsto 101 \quad 110 \mapsto 110 \quad 111 \mapsto 110$$

For arbitrary initial correlation among the memory elements before the computation, the change in memory entropy is:

$$\Delta \mathrm{H}(\mathcal{M}_t)$$
$$= \mathrm{H}(\mathcal{M}_\tau^{(\mathrm{in}_1)}, \mathcal{M}_\tau^{(\mathrm{in}_2)}, \mathcal{M}_\tau^{(\mathrm{out})}) - \mathrm{H}(\mathcal{M}_0^{(\mathrm{in}_1)}, \mathcal{M}_0^{(\mathrm{in}_2)}, \mathcal{M}_0^{(\mathrm{out})})$$
$$\approx \mathrm{H}(\mathcal{M}_0^{(\mathrm{in}_1)}, \mathcal{M}_0^{(\mathrm{in}_2)}, \mathcal{M}_\tau^{(\mathrm{out})}) - \mathrm{H}(\mathcal{M}_0^{(\mathrm{in}_1)}, \mathcal{M}_0^{(\mathrm{in}_2)}, \mathcal{M}_0^{(\mathrm{out})})$$
$$\approx \mathrm{H}(\mathcal{M}_0^{(\mathrm{in}_1)}, \mathcal{M}_0^{(\mathrm{in}_2)}) - \mathrm{H}(\mathcal{M}_0^{(\mathrm{in}_1)}, \mathcal{M}_0^{(\mathrm{in}_2)}, \mathcal{M}_0^{(\mathrm{out})})$$
$$= -\mathrm{H}(\mathcal{M}_0^{(\mathrm{out})} | \mathcal{M}_0^{(\mathrm{in}_1)}, \mathcal{M}_0^{(\mathrm{in}_2)}) ,$$
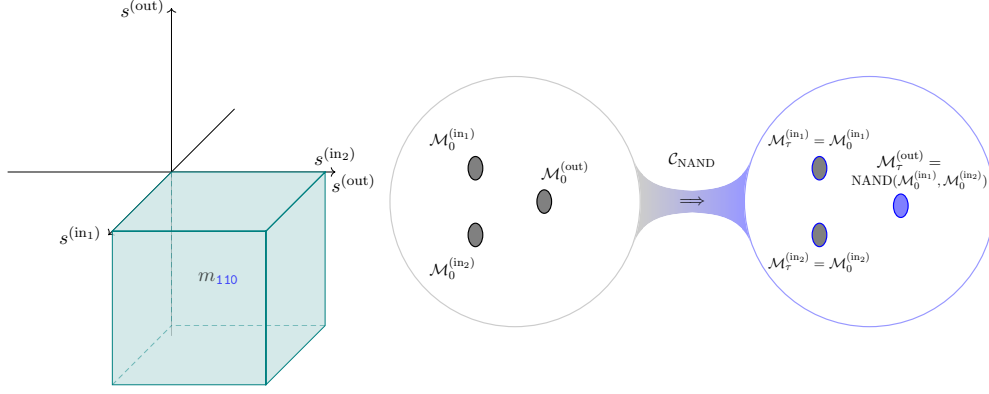
FIG. S1. Composite state space, memory states, and computation associated with the logical NAND operation. [Reproduced with permission (with minor modification) from Ref. [5].]

where we made the approximation that $\epsilon$ is very small. The entropy $\mathrm{H}\big(\mathcal{M}_0^{(\mathrm{out})}\big|\mathcal{M}_0^{(\mathrm{in}_1)}, \mathcal{M}_0^{(\mathrm{in}_2)}\big)$ is the Landauer bound for the NAND gate—a result of interest in its own right. Here, we use it to provide the $\Delta\,\mathrm{H}(\mathcal{M}_t)$ needed for our calculation of dissipation for time-symmetric implementations of NAND.

It should be noted that, by the definition of $\mathcal{C}_{\mathrm{NAND}}$, half of the memory states map to themselves. The other half never get mapped back to themselves upon iteration of the computation. The reciprocity coefficient is therefore:

$$\big\langle \big[\!\big[\mathcal{C}_{\mathrm{NAND}}(\mathcal{C}_{\mathrm{NAND}}(\mathcal{M}_0)) \neq \mathcal{M}_0\big]\!\big]\big\rangle_{\mathcal{M}_0}$$
$$= \boldsymbol{\mu}_0(000) + \boldsymbol{\mu}_0(010) + \boldsymbol{\mu}_0(100) + \boldsymbol{\mu}_0(111)\,.$$

Together with Eq. (29), these give the general lower bound for NAND gate dissipation when a time-symmetric protocol is used and the memory elements are initially correlated.

In the simpler case where the memory elements are initiated with individual biases towards 0 or 1, but are otherwise uncorrelated with each other, the NAND dissipation somewhat simplifies, since then:

$$\Delta\,\mathrm{H}(\mathcal{M}_t) \approx -\,\mathrm{H}\big(\mathcal{M}_0^{(\mathrm{out})}\big|\mathcal{M}_0^{(\mathrm{in}_1)}, \mathcal{M}_0^{(\mathrm{in}_2)}\big)$$
$$= -\,\mathrm{H}\big(\mathcal{M}_0^{(\mathrm{out})}\big)$$
$$= -\,\boldsymbol{H}_{\mathrm{b}}(b_0^{(\mathrm{out})})\,,$$

where $b_0^{(\mathrm{out})}$ is the initial bias of the output bit to be overwritten.

In this latter uncorrelated case, the time-symmetric NAND dissipation resembles the dissipation that would be expected simply from bit erasure of the output bit. The primary difference is that the probability of overwriting the output bit now depends intricately on a function $\big\langle\big[\!\big[\mathcal{C}_{\mathrm{NAND}}(\mathcal{C}_{\mathrm{NAND}}(\mathcal{M}_0)) \neq \mathcal{M}_0\big]\!\big]\big\rangle_{\mathcal{M}_0}$ of the distribution over the joint memory state of all memory elements which is not separable even when the joint distribution over memory elements is separable.

It should be noted that, even in the case of biased but uncorrelated initialization of the memory, the initialization statistics would need to be taken into account to make the time-symmetric physical implementation as thermodynamically efficient as possible. Interestingly, even in the case of time-asymmetric protocols, it has been found that the initialization statistics must be taken into account for thermodynamic efficiency of the NAND gate [5]. Thus, the impetus to match the input statistics is somewhat of a separate generic issue in addition to the lesson of extra dissipation that generically accompanies time-symmetric implementation of reliable computation.