Editors' Suggestion

Finite-Time Landauer Principle

Karel Proesmans,[1,2,*] Jannik Ehrich,[1] and John Bechhoefer[1]
[1]Department of Physics, Simon Fraser University, Burnaby, British Columbia,V5A 1S6, Canada
[2]Hasselt University, B-3590 Diepenbeek, Belgium

(Received 5 June 2020; accepted 14 August 2020; published 3 September 2020)

We study the thermodynamic cost associated with the erasure of one bit of information over a finite amount of time. We present a general framework for minimizing the average work required when full control of a system's microstates is possible. In addition to exact numerical results, we find simple bounds proportional to the variance of the microscopic distribution associated with the state of the bit. In the short-time limit, we get a closed expression for the minimum average amount of work needed to erase a bit. The average work associated with the optimal protocol can be up to a factor of 4 smaller relative to protocols constrained to end in local equilibrium. Assessing prior experimental and numerical results based on heuristic protocols, we find that our bounds often dissipate an order of magnitude less energy.

DOI: 10.1103/PhysRevLett.125.100602

Introduction.—Efficient computation hinges on the ability to erase a memory at minimal energetic cost. The minimum amount of work needed to complete this process is given by the *Landauer limit* [1] stating that at least $k_B T \ln 2$ of work is needed to erase a one-bit memory. Here, k_B is the Boltzmann constant, and T is the absolute temperature at which the erasure process takes place. This work is subsequently released as heat into the environment. Although Landauer's result is a cornerstone in the *thermodynamics of information* [2] and was the key to resolving the paradox around *Maxwell's demon* [3,4], it is achieved only for slow, quasistatic bit erasure. But practical information processing requires fast erasure.

Over the past decade, several experiments have studied the thermodynamics of slow bit erasure and have shown that one can indeed saturate the Landauer bound in the quasistatic limit [5–9]. Those works, along with several theoretical studies [10–12], have suggested that the minimum amount of work needed to erase a bit over a finite amount of time is given by the Landauer limit plus a dissipative correction inversely proportional to the duration of the protocol. The associated proportionality constant, however, depends on the dynamics of the system and the constraints that one puts on the driving protocol. Different protocols lead to different proportionality constants, raising the question of how to select the optimal protocol that minimizes this constant and, hence, the costs of finite-time bit erasure.

Within the fields of *finite-time thermodynamics* [13] and *stochastic thermodynamics* [14,15], the search for protocols that minimize the average dissipation of a mesoscopic thermodynamic system during finite-time transformations has focused on the optimization of a finite (and usually small) number of control parameters influencing the potential landscape of the system [16–21]. For bit erasure, limiting control to a fixed set of parameters may make it more costly or even impossible to fully erase a bit [22–24].

An important advance is the work of Aurell *et al.* [25,26], which uses full control over the potential landscape to find protocols valid in both slow and fast limits that minimize entropy production for a final state constrained to a fixed microscopic probability distribution. However, the need to specify the final distribution is also a limitation, as the entropy production might conceivably be reduced by a different (unknown) choice of final state.

Here, we introduce a framework that uses full control of a potential to achieve efficient, fast bit erasure without knowing in advance the "best" final state. Using this framework, we derive lower and upper bounds on work dissipated during optimal bit erasure. The bounds are proportional to the initial microscopic variance of the bit and confirm that the minimum entropy production is inversely proportional to protocol duration. We also show how to calculate the minimum work required for a given potential shape and given erasure time. Compared to previous experimental and numerical studies, taking advantage of full potential control can reduce the cost of fast erasure by roughly an order of magnitude.

In an accompanying paper [27], we give full details of the calculations and generalize to the case of *partial erasure* of information in the bit.

Thermodynamic cost of finite-time transformations.—Consider a bit encoded in a system described by a microscopic variable x. The bit is in state 1 if $x > 0$ and state 0 if $x < 0$. The probability density $p(x,t)$ of x is described by a Fokker-Planck equation:

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(p(x,t)\frac{\partial}{\partial x}V(x,t)\right) + \frac{\partial^2}{\partial x^2}p(x,t), \quad (1)$$

where $V(x,t)$ is the potential energy landscape. In Eq. (1), we have scaled energy by k_BT and lengths by $x_0 \equiv \sqrt{\mathrm{Var}(x)}$, the variance of the equilibrium distribution for the potential $V(x,0) \equiv V_0(x)$. Time is scaled by x_0^2/D, with D the diffusion coefficient. This description applies to a broad class of systems, including colloidal particles trapped in a potential [5–8] and superconducting fluxes [9]. In such systems, the microscopic state $x(t)$ is coarse grained to two (or more) macrostates that encode information [1].

Building on ideas from stochastic thermodynamics [14,15] and optimal-transport theory [28,29], one can calculate the minimum average work to go from an initial microscopic equilibrium distribution $p_0(x)$ to a final microscopic distribution $p_\tau(x)$ over a time interval of length τ for protocols having the same start and end point, with $V(x,0) = V(x,\tau) \equiv V_0(x)$. Assuming full control over the potential $V(x,t)$, one finds [25,26,30,31]

$$W = \underbrace{\int_{-\infty}^{\infty} dx\, p_\tau(x) \ln \frac{p_\tau(x)}{p_0(x)}}_{\Delta\mathcal{F}} + \underbrace{\frac{1}{\tau}\int_0^1 dy\, [f_0^{-1}(y) - f_\tau^{-1}(y)]^2}_{\Delta_i S}, \tag{2}$$

where $f_{0/\tau}(x) = \int_{-\infty}^{x} dx' p_{0/\tau}(x')$ are the associated cumulative distributions and f^{-1} their inverses. The quantity $\Delta\mathcal{F}$ is the change in nonequilibrium free energy arising solely because the probability density is transformed from $p_0(x)$ to $p_\tau(x)$, and $\Delta_i S$ is the average entropy production of the transformation [32]. A similar expression holds for discrete systems [33,34]. After the protocol has completed, the system relaxes back into a local equilibrium state, so that all work is released as heat into the environment.

For many practical applications, one is interested not in the exact "microscopic" distribution of the system but rather in a coarse-grained, "macroscopic" distribution. For example, when erasing a bit of memory, one is not interested in the full distribution of the microscopic variable but only in the probability for the bit to be in macrostate 0, $P_L = p(x<0)$, or in macrostate 1, $P_R = p(x>0)$. This means that fully minimizing the amount of work to go from an initial macroscopic distribution to a different final macroscopic distribution implies a *second* minimization of Eq. (2) over all possible microscopic distributions $p_\tau(x)$ that are compatible with the desired final macroscopic distribution. [The initial distribution $p_0(x)$ is fixed if we start in thermal equilibrium.] Following Refs. [25,26,30,31], we change variables from $f_\tau(x)$ to $\Gamma(x) \equiv f_\tau^{-1}[f_0(x)]$. The minimum amount of work required to complete the process is then

$$W_{\min} = \min_{\Gamma(x)} \int_{-\infty}^{\infty} dx\, p_0(x) \times \left[\ln \frac{p_0(x)}{\Gamma'(x)\, p_0[\Gamma(x)]} + \frac{[\Gamma(x)-x]^2}{\tau}\right], \tag{3}$$

where the minimization is done over all $\Gamma(x)$ that correspond to the correct macroscopic final distribution. For "full" bit erasure at time τ, the particle is always somewhere within the region corresponding to macrostate 0 ($x<0$), so that $P_L = 1$ and $P_R = 0$ for $t = \tau$. Consequently,

$$f_\tau(0) = 1, \quad \text{or} \quad \Gamma[f_0^{-1}(1)] = 0, \tag{4}$$

where the first condition again implies that all probability density is in $x<0$ and the second implies that, at $t=0$, we have $f_0(\infty) = 1$ and, hence, $\Gamma(\infty) = 0$, since Γ maps positions at time τ to positions at $t=0$. For boundary conditions appropriate to partial erasure, see Ref. [27].

Using the calculus of variations, we find that the optimal $\Gamma(x)$, and, therefore, the optimal final microscopic distribution, obeys the Euler-Lagrange equation [27]

$$V'[\Gamma(x)] - \frac{V'(x)}{\Gamma'(x)} - \frac{\Gamma''(x)}{\Gamma'(x)^2} + \frac{2}{\tau}[\Gamma(x) - x] = 0, \tag{5}$$

where we have assumed that, at $t=0$, the system is in equilibrium, $p_0(x) \sim \exp[-V_0(x)]$.

Bounds on finite-time bit erasure.—Having formulated a general theory, we apply it to the problem of bit erasure. We consider a system described by Eq. (1), with a potential that is initially symmetric, $V_0(x) = V_0(-x)$. Although solving the minimization condition [Eq. (5)] cannot, in general, be done analytically, we can nonetheless place upper and lower bounds on $W_{\min}$.

To establish an upper bound for $W_{\min}$, we fix the final microscopic distribution to be the local-equilibrium distribution that fixes all probability to be in the region $x<0$ (Fig. 1, bottom right):

$$p_\tau(x) = p_{\mathrm{leq}}(x) = \begin{cases} 2p_0(x), & x<0, \\ 0, & x>0. \end{cases} \tag{6}$$

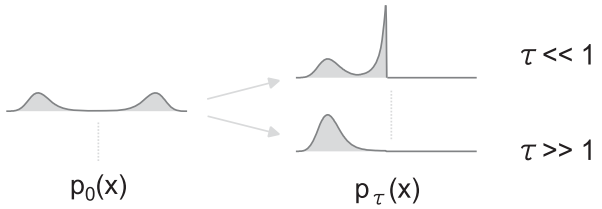

FIG. 1. Optimal protocols in the short- and long-time limits. Starting from the equilibrium distribution $p_0(x)$ for a symmetric potential $V_0(x)$ at time $t=0$ (left distribution), the system state is transformed to $p_\tau(x)$ at time τ. For $\tau \ll 1$ (upper right distribution), the final distribution is approximately the sum of the initial probability density of the left well plus a sharp peak composed of probability transported from the right well. For $\tau \gg 1$ (lower right distribution), the final distribution is in local equilibrium in the left well.

The local equilibrium distribution minimizes the first term in Eq. (3), in accordance with the boundary condition of full erasure.

The optimal protocol for this case leads to an average work $W_{\mathrm{min,leq}}$. We have $W_{\mathrm{min}} \leq W_{\mathrm{min,leq}}$, because constraining the final distribution to a local equilibrium can only increase the work relative to the case where we allow the final distribution to be selected from a set of distributions with $p_\tau(x > 0) \equiv 0$. In Ref. [27], we show that

$$W_{\mathrm{min,leq}} \leq \ln 2 + \frac{2}{\tau}. \tag{7}$$

In Ref. [27], we also derive an alternate lower bound for $W_{\mathrm{min,leq}}$ based on Ref. [35].

To derive a lower bound on W_{min}, we observe that Eq. (3) minimizes the sum of two terms. Minimizing each term separately then gives a lower bound:

$$\begin{aligned} W_{\mathrm{min}} &\geq \min_{\Gamma(x)} \int_{-\infty}^{\infty} dx p_0(x) \ln \frac{p_0(x)}{\Gamma'(x) p_0[\Gamma(x)]} + \min_{\Gamma(x)} \int_{-\infty}^{\infty} dx p_0(x) \frac{[\Gamma(x) - x]^2}{\tau} \\ &= \min_{p_\tau(x)} \int_{-\infty}^{\infty} dx p_\tau(x) \ln \frac{p_\tau(x)}{p_0(x)} + \frac{1}{\tau} \int_0^{\infty} dx p_0(x) x^2 \\ &= \ln 2 + \frac{1}{2\tau}. \end{aligned} \tag{8}$$

As with the upper bound, the first term of Eq. (8) is also minimized by the local-equilibrium distribution. By contrast, the optimal choice of Γ in the second term is $\Gamma(x) = x$ for $x < 0$ (which minimizes the integrand) and $= 0$ for $x \geq 0$ (because no probability is left at the end for $x \geq 0$). More visually, the optimal protocol for the second term "pushes" the probability initially in the right well to a spike at $x \approx 0^-$. The probability in the "wrong" well is moved the minimum amount possible to be in the correct macrostate—pushed to its edge—while the probability already in the macrostate is left untouched (Fig. 1, top right). As $\tau \to 0$, the spike of probability at the edge of the macrostate approaches a delta function.

Piling the probability into a delta function leads to an infinite contribution from the first term, since the free energy of a perfectly localized particle is infinite; however, in the limit $\tau \to 0$, the second term has an infinite weight $\tau^{-1} \to \infty$, allowing for such singular behavior.

As the main result of this paper, we rewrite these bounds in terms of the original unscaled quantities:

$$\ln 2 + \frac{\mathrm{Var}(x)}{2D\tau} \leq \frac{W_{\mathrm{min}}}{k_B T} \leq \frac{W_{\mathrm{min,leq}}}{k_B T} \leq \ln 2 + \frac{2\mathrm{Var}(x)}{D\tau}. \tag{9}$$

That is, the cost to fully erase a bit over a finite amount of time is equal to the Landauer cost ln 2 plus a term that is determined by the initial variance of the distribution. Remarkably, for all τ, the minimum entropy production is always $\sim\tau^{-1}$. We notice, in particular, that the upper and lower bounds to the entropy production differ by a factor of 4. We can understand this numerical factor by noting that an approximate expression for the dissipation is

$$\frac{F\Delta x}{k_B T} \sim \frac{\gamma}{k_B T} \left(\frac{\Delta x}{\tau}\right) \Delta x \sim \frac{(\Delta x)^2}{D\tau}, \tag{10}$$

where we apply the friction force $F = \gamma \dot{x} \approx \gamma(\Delta x/\tau)$ and the Einstein relation $D = k_B T/\gamma$. The quantity Δx is the typical distance a particle is transported during the protocol. In the long-time limit, the system stays in local equilibrium, and the probability from the right well is shifted to the left well. In the short-time limit, the same probability is moved only half as far (by symmetry of the potential) to $x = 0$. The factor-2 reduction in Δx decreases the dissipation by a factor of 4.

In special cases, we can saturate the bounds in Eq. (9). In the fast-erasure limit $\tau \ll 1$, the τ^{-1} term in Eq. (8) dominates, leading to saturation at the lower bound in Eq. (9), giving a general, closed expression for the cost of fast erasure of a bit. In the slow-erasure limit $\tau \gg 1$, the erasure cost reduces to the Landauer cost, as expected. For a first-order correction in $1/\tau$, one can verify that $W_{\mathrm{min}} = W_{\mathrm{min,leq}}$, saturating the second inequality in Eq. (9) [27]. Finally, the last inequality in Eq. (9) is saturated for an initial "two-state" distribution:

$$p_0(x) = \frac{1}{2}\left[\delta\left(x - \frac{1}{2}\Delta x\right) + \delta\left(x + \frac{1}{2}\Delta x\right)\right], \tag{11}$$

where Δx is the difference in x between the two states. Equation (11) is the limiting distribution for a broad class of double-well potentials with an infinite barrier between the wells.

Example.—Let the initial energy landscape be (Fig. 2)

$$V_0(x) = E_b \left[\left(\frac{x}{x_m}\right)^2 - 1\right]^2, \tag{12}$$

with a barrier $E_b = 4$ between the wells and $x_m \approx 1.04$, which implies an equilibrium variance $\mathrm{Var}(x) = 1$. Figure 2 shows the optimal protocol $V(x,t)$ and

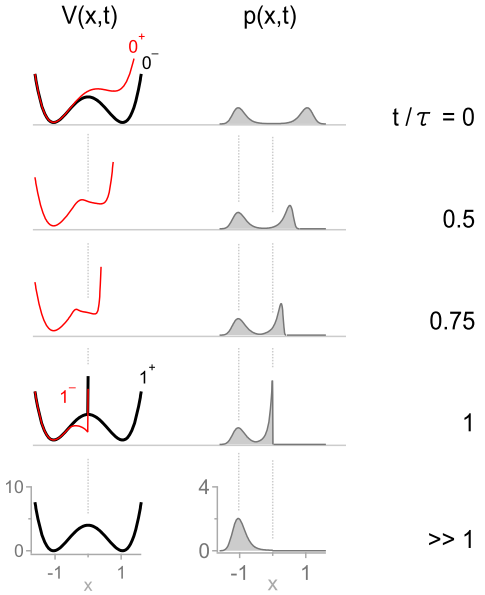

FIG. 2. Erasure protocol for $\tau = 0.2$ and $E_b = 4$. The original and final double-well potentials $V_0(x)$ are shown in black at the top and bottom of the left column. Red potentials $V(x,t)$ denote the control. The control is carried out for $0^+ < t/\tau < 1^-$, and the potential changes discontinuously at $t = 0$ and $t/\tau = 1$. At $t/\tau = 1$, the probability distribution is peaked near $x = 0$. Therefore, an infinitesimally narrow, extra barrier $\delta(0)$ is added to $V_0(x)$ to prevent probability from leaking back into the right well. It is removed at a later time $t/\tau \gg 1$, after the system has relaxed to local equilibrium and the peak in the probability distribution has disappeared. The right-hand column shows corresponding probability distributions.

corresponding densities $p(x,t)$ for $\tau = 0.2$. Once the optimal $\Gamma(x)$ is determined, the intermediate probability distribution $p(x,t)$ can be computed, which gives the intermediate control protocol $V(x,t)$ by inverting the Fokker-Planck equation [Eq. (1)]. See Ref. [27], Sec. II for full details. The protocol has jump discontinuities when passing from $V_0(x)$ (black curve) at $t = 0^-$ to $V(x, t = 0^+)$ (red curve) and similarly in passing from $V(x, t = 1^-)$ to $V_1(x) = V_0(x)$. At $t = \tau$, we add a $\delta(0)$ barrier that keeps probability from leaking back into the right well for $t > \tau$. Notice that no work is done for $t > \tau$. The probability that is trapped in the left well then relaxes to local equilibrium, after which the barrier may be removed. See the bottom plots.

We then numerically calculate the upper and lower bounds [Eqs. (7) and (8)] and $W_{\min}$ and $W_{\min,\mathrm{leq}}$ for full erasure. Figure 3 shows that the upper and lower bounds are satisfied. We also note that $W_{\min} \approx W_{\min,\mathrm{leq}}$ in the slow-driving limit and that $W_{\min}$ saturates the lower bound in the fast-driving limit. For $E_b \gg 1$, the potential wells are quite steep, and the Boltzmann distribution resembles quite well the two-delta-function distribution [Eq. (11)], which explains why $W_{\min,\mathrm{leq}}$ is close to the upper bound.

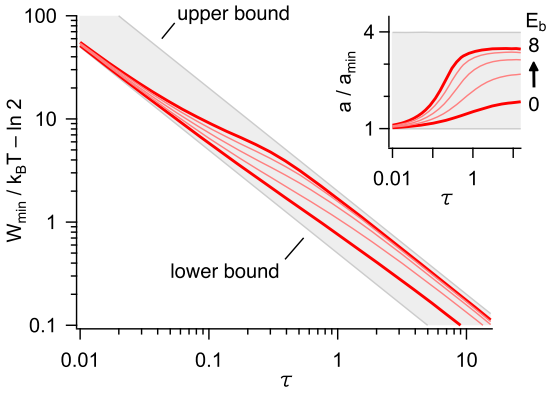

FIG. 3. Minimum entropy production in excess of the Landauer bound. The shaded regions show upper $(2\tau^{-1})$ and lower $(\tau^{-1}/2)$ bounds. The inset shows $a \equiv \tau(W_{\min}/k_B T - \ln 2)$ relative to its lower bound. Red curves are plotted for $E_b = \{0, 2, 4, 6, 8\}$. Heavier lines denote the $E_b = \{0, 8\}$ cases.

Comparison with experimental and numerical results.— Over the past decade, several high-precision tests of the Landauer principle have been performed [5,6,8,36,37]. In general, those protocols satisfied $W/k_B T - \ln 2 \sim \tau^{-1}$ in the slow-driving limit. Therefore, for large τ, the measured work in those experimental protocols has qualitatively the same form as the optimal protocol, raising the question of how close the experimental results are to the optimum. From Table I, we can see that the measured amount of entropy production exceeds the optimum by factors of 2–6 (see Supplemental Material [38]).

Furthermore, we can also compare our bound to numerical studies of bit erasure. Zulkowski and DeWeese [10] calculate the amount of work to erase a bit in a potential consisting of two flat wells of length ℓ, separated by a thin wall of arbitrary height. If one controls only the height of the wells and uses slow driving, they showed that the minimum amount of work to erase a bit is given by $D\tau/\mathrm{Var}(x)(W/k_B T - \ln 2) = 3[(\sqrt{2}-1)^2 + 1] \approx 3.51$. By contrast, Boyd *et al.* [12] derived a general framework to calculate the work to erase a bit via a protocol that keeps the system always in local equilibrium. For the flat-well potential used by Zulkowski and DeWeese, this protocol actually performs better than the limited-control protocol used in Ref. [10]: $D\tau/\mathrm{Var}(x)(W/k_B T - \ln 2) = \pi^2(2-\sqrt{2})/2 \approx 2.89$. However, for a double-well potential

TABLE I. Comparison between work measurements found in the literature and $W_{\min}$ in terms of $\mathrm{Var}(x)/D\tau$ [38].

Experiment/numerics	$W/k_B T - \ln 2$	$W_{\min}/k_B T - \ln 2$	Ratio
Bérut *et al.* [5,36,37]	10.2	1.80	5.67
Gavrilov *et al.* [8]	7.20	1.82	3.96
Jun *et al.* [6]	5.67	1.82	3.11
Zulkowski *et al.* [10]	3.51	1	3.51
Boyd *et al.* [12]	2.89	1	2.89

of the form of Eq. (12) and $E_b = 10$, applying the method of Ref. [12] leads to average work values that are several orders of magnitude larger.

All the above protocols were explored in the slow ($\tau \to \infty$) limit. But we have shown here that the entropy production for optimal protocols, when scaled by $\mathrm{Var}(x)/D\tau$, drops to $1/2$ for fast driving ($\tau \to 0$). Thus, for fast erasure, our protocol can improve efficiency by up to a further factor of 4.

Conclusions and outlook.—When erasing a bit, dissipation is minimized by moving probability as little as possible, given the final-state constraint. Long protocols are automatically in local equilibrium, but short protocols can increase performance by moving probability to the edge of the desired macrostate. In one dimension, the move is half the distance compared to one that maintains local equilibrium, reducing dissipation by up to a factor of 4.

We suggest three extensions of our formalism: (i) *Higher dimensions.*—The factor-4 improvement results from the one-dimensional geometry. Although the extension to higher-dimensional systems is nontrivial as the dissipation depends on the Wasserstein distance between the initial and final distribution, which generally cannot be calculated analytically for higher-dimensional systems [28], it might be possible to use symmetry properties to extend our result to higher-dimensional systems. (ii) *Reduced damping.*—Bit erasure might be more efficient for critically damped systems under limited control [39]. On the other hand, the minimum amount of dissipation needed to transform an overdamped system from an initial to a final state is also a lower bound for the amount of dissipation needed to perform the same transformation in underdamped systems [40]. Therefore, we expect our lower bound [Eq. (8)] to also hold for underdamped systems. (ii) *Quantum effects.*—Landauer's principle also holds for quantum systems strongly interacting with their environment [40,41]. Furthermore, stronger bounds have recently been derived to include quantum-mechanical effects [42,43]. It would be interesting to see whether similar quantum corrections can be calculated for Eq. (9).

A remarkable feature of the optimal solutions is the existence of various discontinuities and singularities in the control. Here, as elsewhere [16], there are discontinuities in the potential at the beginning and end of the protocol; in addition, the intermediate-time potential can have near discontinuities in the slope [26], which become more pronounced for fast driving (Fig. 2). Unfortunately, experimental systems are likely unable to reproduce such protocols exactly [44,45]. Moreover, optimal protocols assume a "perfect" model of the system under control. But parameters are always uncertain, and the shape of the underlying potential $V_0(x)$ may simplify a more complex reality. For such reasons, experiments can only approximate the optimal solutions derived here. The challenge—and this is what makes optimal control a problem of physics as well as mathematics—is to find good approximations to the "best" control [46] that are robust to imperfections of system models and to experimental constraints.

We thank David Sivak and Raphaël Chétrite for helpful comments. This work was supported by a Foundational Questions Institute Grant No. FQXi-RFP-2019-IAF, and by a National Science and Engineering Research Council of Canada Discovery Grant.

*Karel_Proesmans@sfu.ca

[1] R. Landauer, Irreversibility and heat generation in the computing process, IBM J. Res. Dev. **5**, 183 (1961).
[2] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, Thermodynamics of information, Nat. Phys. **11**, 131 (2015).
[3] *Maxwell's Demon: Entropy, Information, Computing*, edited by H. S. Leff and A. F. Rex (Princeton University Press, Princeton, NJ, 1990).
[4] A. Rex, Maxwell's demon—A historical review, Entropy **19**, 240 (2017).
[5] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, Experimental verification of Landauer's principle linking information and thermodynamics, Nature (London) **483**, 187 (2012).
[6] Y. Jun, M. Gavrilov, and J. Bechhoefer, High-Precision Test of Landauers Principle in a Feedback Trap, Phys. Rev. Lett. **113**, 190601 (2014).
[7] M. Gavrilov and J. Bechhoefer, Erasure without Work in an Asymmetric Double-Well Potential, Phys. Rev. Lett. **117**, 200601 (2016).
[8] M. Gavrilov, R. Chétrite, and J. Bechhoefer, Direct measurement of weakly nonequilibrium system entropy is consistent with Gibbs-Shannon form, Proc. Natl. Acad. Sci. U.S.A. **114**, 11097 (2017).
[9] O.-P. Saira, M. H. Matheny, R. Katti, W. Fon, G. Wimsatt, J. P. Crutchfield, S. Han, and M. L. Roukes, Nonequilibrium thermodynamics of erasure with superconducting flux logic, Phys. Rev. Research **2**, 013249 (2020).
[10] P. R. Zulkowski and M. R. DeWeese, Optimal finite-time erasure of a classical bit, Phys. Rev. E **89**, 052140 (2014).
[11] P. R. Zulkowski and M. R. DeWeese, Optimal control of overdamped systems, Phys. Rev. E **92**, 032117 (2015).
[12] A. B. Boyd, A. Patra, C. Jarzynski, and J. P. Crutchfield, Shortcuts to thermodynamic computing: The cost of fast and faithful erasure, arXiv:1812.11241.
[13] B. Andresen, Current trends in finite-time thermodynamics, Angew. Chem. Int. Ed. **50**, 2690 (2011).
[14] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, Rep. Prog. Phys. **75**, 126001 (2012).
[15] C. Van den Broeck and M. Esposito, Ensemble and trajectory thermodynamics: A brief introduction, Physica (Amsterdam) **418A**, 6 (2015).
[16] T. Schmiedl and U. Seifert, Optimal Finite-Time Processes in Stochastic Thermodynamics, Phys. Rev. Lett. **98**, 108301 (2007).

[17] M. V. Bonança and S. Deffner, Optimal driving of isothermal processes close to equilibrium, J. Chem. Phys. **140**, 244119 (2014).
[18] D. A. Sivak and G. E. Crooks, Thermodynamic Metrics and Optimal Paths, Phys. Rev. Lett. **108**, 190602 (2012).
[19] S. Tafoya, S. J. Large, S. Liu, C. Bustamante, and D. A. Sivak, Using a system's equilibrium behavior to reduce its energy dissipation in nonequilibrium processes, Proc. Natl. Acad. Sci. U.S.A. **116**, 5920 (2019).
[20] C. A. Plata, D. Guéry-Odelin, E. Trizac, and A. Prados, Finite-time adiabatic processes: Derivation and speed limit, Phys. Rev. E **101**, 032129 (2020).
[21] S. J. Bryant and B. B. Machta, Energy dissipation bounds for autonomous thermodynamic cycles, Proc. Natl. Acad. Sci. U.S.A. **117**, 3478 (2020).
[22] G. Diana, G. B. Bagci, and M. Esposito, Finite-time erasing of information stored in fermionic bits, Phys. Rev. E **87**, 012111 (2013).
[23] A. B. Boyd, D. Mandal, and J. P. Crutchfield, Thermodynamics of modularity: Structural costs beyond the Landauer bound, Phys. Rev. X **8**, 031036 (2018).
[24] P. M. Riechers, A. B. Boyd, G. W. Wimsatt, and J. P. Crutchfield, Balancing error and dissipation in highly-reliable computing, arXiv:1909.06650.
[25] E. Aurell, C. Mejía-Monasterio, and P. Muratore-Ginanneschi, Optimal Protocols and Optimal Transport in Stochastic Thermodynamics, Phys. Rev. Lett. **106**, 250601 (2011).
[26] E. Aurell, K. Gawedzki, C. Mejia-Monasterio, R. Mohayaee, and P. Muratore-Ginanneschi, Refined second law of thermodynamics for fast random processes, J. Stat. Phys. **147**, 487 (2012).
[27] K. Proesmans, J. Ehrich, and J. Bechhoefer, companion paper, Optimal finite-time bit erasure under full control, Phys. Rev. E **102**, 032105 (2020).
[28] C. Villani, *Topics in Optimal Transportation* (American Mathematical Society, Providence, 2003), Vol. 58.
[29] B. Lévy and E. L. Schwindt, Notions of optimal transport theory and how to implement them on a computer, Comput. Graph. **72**, 135 (2018).
[30] Y. Zhang, Work needed to drive a thermodynamic system between two distributions, Europhys. Lett. **128**, 30002 (2019).
[31] Y. Zhang, Optimization of stochastic thermodynamic machines, J. Stat. Phys. **178**, 1336 (2020).
[32] M. Esposito and C. Van den Broeck, Second law and Landauer principle far from equilibrium, Europhys. Lett. **95**, 40004 (2011).
[33] N. Shiraishi, K. Funo, and K. Saito, Speed Limit for Classical Stochastic Processes, Phys. Rev. Lett. **121**, 070601 (2018).
[34] Y. Zhang, Comment on "Speed limit for classical stochastic processes", arXiv:1811.06978.
[35] A. Dechant and Y. Sakurai, Thermodynamic interpretation of Wasserstein distance, arXiv:1912.08405.
[36] A. Bérut, A. Petrosyan, and S. Ciliberto, Detailed Jarzynski equality applied to a logically irreversible procedure, Europhys. Lett. **103**, 60002 (2013).
[37] A. Bérut, A. Petrosyan, and S. Ciliberto, Information and thermodynamics: Experimental verification of Landauer's Erasure principle, J. Stat. Mech. (2015) P06015.
[38] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.125.100602 for discussion of experimental setups and parameter values, along with accompanying calculations on the various bounds.
[39] A. Deshpande, M. Gopalkrishnan, T. E. Ouldridge, and N. S. Jones, Designing the optimal bit: Balancing energetic cost, speed and reliability, Proc. R. Soc. A **473**, 20170117 (2017).
[40] D. Reeb and M. M. Wolf, An improved Landauer principle with finite-size corrections, New J. Phys. **16**, 103011 (2014).
[41] M. H. Mohammady, M. Mohseni, and Y. Omar, Minimising the heat dissipation of quantum information erasure, New J. Phys. **18**, 015011 (2016).
[42] A. M. Timpanaro, J. P. Santos, and G. T. Landi, Landauers Principle at Zero Temperature, Phys. Rev. Lett. **124**, 240601 (2020).
[43] H. J. Miller, G. Guarnieri, M. T. Mitchison, and J. Goold, Quantum fluctuations hinder finite-time information erasure near the Landauer limit, arXiv:2007.01882.
[44] I. A. Martínez, A. Petrosyan, D. Guéry-Odelin, E. Trizac, and S. Ciliberto, Engineered swift equilibration of a Brownian particle, Nat. Phys. **12**, 843 (2016).
[45] M. Chupeau, B. Besga, D. Guéry-Odelin, E. Trizac, A. Petrosyan, and S. Ciliberto, Thermal bath engineering for swift equilibration, Phys. Rev. E **98**, 010104(R) (2018).
[46] T. R. Gingrich, G. M. Rotskoff, G. E. Crooks, and P. L. Geissler, Near-optimal protocols in complex nonequilibrium transformations, Proc. Natl. Acad. Sci. U.S.A. **113**, 10263 (2016).