

Optimal finite-time bit erasure under full control

Karel Proesmans,^{1,2,*} Jannik Ehrich¹ and John Bechhoefer¹

¹*Department of Physics, Simon Fraser University, Burnaby, B.C., V5A 1S6, Canada*

²*Hasselt University, B-3590 Diepenbeek, Belgium*

 (Received 5 June 2020; accepted 11 August 2020; published 3 September 2020)

We study the finite-time erasure of a one-bit memory consisting of a one-dimensional double-well potential, with each well encoding a memory macrostate. We focus on setups that provide full control over the form of the potential-energy landscape and derive protocols that minimize the average work needed to erase the bit over a fixed amount of time. We allow for cases where only some of the information encoded in the bit is erased. For systems required to end up in a local-equilibrium state, we calculate the minimum amount of work needed to erase a bit explicitly, in terms of the equilibrium Boltzmann distribution corresponding to the system's initial potential. The minimum work is inversely proportional to the duration of the protocol. The erasure cost may be further reduced by relaxing the requirement for a local-equilibrium final state and allowing for any final distribution compatible with constraints on the probability to be in each memory macrostate. We also derive upper and lower bounds on the erasure cost.

DOI: [10.1103/PhysRevE.102.032105](https://doi.org/10.1103/PhysRevE.102.032105)

I. INTRODUCTION

In his seminal 1961 paper, Landauer showed that a finite amount of work, known as the *Landauer bound*, is needed to erase a bit of information [1,2]. This bound has been verified in several theoretical models [3–5] and experimental tests [6–11]. However, the Landauer bound can be saturated only in the quasistatic limit, an assumption that is never valid for practical systems.

How much extra work, then, is needed to erase a bit in a finite amount of time? Several theoretical studies have addressed this question using assumptions such as slow driving [12,13], a final distribution that is locally in equilibrium [14], or counterdiabatic driving [15].

In an accompanying paper [16], we have derived a general theory to optimize the work needed to erase a bit over a finite amount of time, provided that one has full control over the external potential. Often, however, one is interested in situations where only part of the information encoded in the bit is erased (*partial erasure*) [5,12,13,15,17]. In this paper, we generalize the results from Ref. [16] to partial erasure, provide full details of the calculations, and derive a new set of upper and lower bounds associated with the finite-time erasure of a bit, while still allowing full control over the potential energy.

In Sec. II, we introduce the physical model of a bit and review the existing literature on optimal transport theory. These results are used in Sec. III to calculate the amount of work needed to erase a bit, provided that the final state is in local equilibrium. Subsequently, in Sec. IV, we relax the assumption on the final distribution. Our results are tested on two simple toy models in Sec. V. Finally, we give a thorough discussion and outlook in Sec. VI.

II. SETUP: OPTIMAL FINITE-TIME BIT ERASURE

Let us consider the work required to transform a system whose state x is described by a probability density function $p_0(x)$ at an initial time 0 to a density $p_\tau(x)$ at a final time τ . At the initial and final times, the potential-energy landscape is $V_0(x)$. Under this assumption, the required work comes purely from the changes in the occupation probabilities of the macrostates. At intermediate times t , the potential may be altered arbitrarily, as specified by the function $V(x, t)$. Discontinuities in $V(x, t)$ are allowed as they correspond to changing the applied force suddenly at the position of the particle, which is an allowable operation.

Throughout this paper, we will assume that the dynamics of the system can be described by an overdamped Fokker-Planck equation,

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

where we have scaled entropy by the Boltzmann constant, k_B , energies by the thermal energy of the environment, $k_B T$, lengths (squared) by $\text{Var}(x)$, the variance of $p_0(x) \equiv p(x, 0)$, and time by $\text{Var}(x)/D$. We denote the cumulative distributions associated with $p_0(x)$ and $p_\tau(x) \equiv p(x, \tau)$ by

$$f_{0/\tau}(x) = \int_{-\infty}^x dx' p_{0/\tau}(x'). \quad (2)$$

Our goal is to find a protocol $V(x, t)$ that minimizes the average work done on the system as it is transformed from $p_0(x)$ to $p_\tau(x)$. In this work, we assume that the system is in global thermodynamic equilibrium initially ($t = 0$). Thus, $p_0(x) \sim \exp(-\beta V_0(x))$, where $\beta = (k_B T)^{-1}$ is the inverse temperature.

The analysis of Refs. [14,18–20] shows that the optimal control protocol leads to the following intermediate

*karel_proesmans@sfu.ca

cumulative distribution (using notation that follows Refs. [19,20]):

$$f(x, t) = f_0(\Gamma_t^{-1}(x)), \quad (3)$$

where

$$\Gamma_t(x) \equiv x + \frac{t}{\tau} [f_\tau^{-1}(f_0(x)) - x]. \quad (4)$$

Applying the coordinate transformation in Eq. (4) leads to the intermediate probability distribution

$$p(x, t) = \frac{p_0(\Gamma_t^{-1}(x))}{\Gamma_t'(\Gamma_t^{-1}(x))}. \quad (5)$$

The potential, or protocol, $V(x, t)$ that achieves this intermediate distribution $p(x, t)$ is determined by inverting the Fokker-Plank equation (1):

$$V(x, t) = -\ln p(x, t) + \int_{-\infty}^x dx' \frac{\int_{-\infty}^{x'} dx'' \frac{\partial}{\partial t} p(x'', t)}{p(x', t)}. \quad (6)$$

When this optimal control potential is chosen, the minimum average work is [14,18–20]

$$\begin{aligned} W_{\min|p_\tau} &= \Delta\mathcal{F} + \Delta_i S \\ &= \int_{-\infty}^{\infty} dx p_\tau(x) \ln \frac{p_\tau(x)}{p_0(x)} \\ &\quad + \frac{1}{\tau} \int_0^1 dy [f_0^{-1}(y) - f_\tau^{-1}(y)]^2, \end{aligned} \quad (7)$$

where the first term, $\Delta\mathcal{F}$, is the change in nonequilibrium free energy [21] and the second term, $\Delta_i S$, the average entropy production [22,23]. Note that the cumulative probability $y = f_0(x)$ is dimensionless. We also stress that the work needed to erase the bit will be dissipated as heat into the environment. Therefore, our bound on the work needed to erase a bit could alternatively be written as a bound on the heat produced while erasing a bit, in agreement with Landauer's original formulation.

In Eq. (7), we denote the average work by $W_{\min|p_\tau}$ to emphasize that the choice is conditioned on $p_\tau(x)$. It also depends on $p_0(x)$, but we always fix $p_0(x)$ to be the equilibrium distribution for the given potential, $V_0(x)$.

We now apply this general framework to the problem of bit erasure by focusing on a bit described by a microscopic variable x , where the bit is in macrostate 1 if $x > 0$ and macrostate 0 if $x < 0$. We assume that the bit is initially in equilibrium with a symmetric potential $V_0(-x) = V_0(x)$, with an equilibrium distribution $p_0(x)$. The bit thus has probability $1/2$ to be either in macrostate 0 or 1.

We allow for the possibility of *partial erasure*, where a fraction $\epsilon \leq 1/2$ of probability, the *erasure error*, remains in the right well at the end of the protocol. Thus, at $t = \tau$, the system is in macrostate 0 with probability $p_\tau(x < 0) = 1 - \epsilon$ and in macrostate 1 with probability $p_\tau(x > 0) = \epsilon$.

III. ERASURE TO LOCAL EQUILIBRIUM

Previous studies of the erasure of a one-bit memory have focused on cases where, at the end of the protocol, the

distribution is locally in equilibrium. In experimental realizations, this condition arises because the manipulations are slow [6–9,11]. The assumed form of the final distribution also simplifies theoretical calculations [12,14,15].

In this section, we generalize such calculations to the case of partial erasure, for which

$$p_{\text{leq}}(x) = \begin{cases} 2(1 - \epsilon) p_0(x) & x < 0 \\ 2\epsilon p_0(x) & x \geq 0. \end{cases} \quad (8)$$

From Eq. (7), the minimum work needed to erase a symmetric potential in finite time with a final distribution equal to the local-equilibrium distribution is

$$\begin{aligned} W_{\min, \text{leq}} &= W_{\min|p_\tau = p_{\text{leq}}} \\ &= \Delta\mathcal{F}_{\text{leq}} + \Delta_i S_{\min, \text{leq}}, \end{aligned} \quad (9)$$

where $\Delta\mathcal{F}_{\text{leq}}$ can be evaluated using Eq. (8). Taking into account the requirements that $p(x, 0) = p_0(x)$ be the equilibrium Boltzmann distribution and that the potential at the end has the same form as at the beginning, $V(x, 0) = V(x, \tau) = V_0(x)$, we have

$$\begin{aligned} \Delta\mathcal{F}_{\text{leq}} &= \int_{-\infty}^{\infty} dx p_{\text{leq}}(x) \ln \frac{p_{\text{leq}}(x)}{p_0(x)} \\ &= \ln 2 + \epsilon \ln \epsilon + (1 - \epsilon) \ln(1 - \epsilon). \end{aligned} \quad (10)$$

Similarly, the entropy production is

$$\Delta_i S_{\min, \text{leq}} = \frac{1}{\tau} \int_0^1 dy [f_0^{-1}(y) - f_{\text{leq}}^{-1}(y)]^2, \quad (11)$$

where $f_{\text{leq}}^{-1}(y)$ is the inverse of the cumulative local-equilibrium distribution $f_{\text{leq}}(x)$.

As we show in Appendix A, a local-equilibrium final distribution [Eq. (8)] implies that the entropy production can be rewritten purely in terms of the initial equilibrium distribution:

$$\begin{aligned} \Delta_i S_{\min, \text{leq}} &= \frac{2}{\tau} \left\{ 1 - \int_0^{1-\epsilon} dy f_0^{-1}(y) f_0^{-1} \left[\frac{y}{2(1-\epsilon)} \right] \right. \\ &\quad \left. - \int_{1-\epsilon}^1 dy f_0^{-1}(y) f_0^{-1} \left(1 + \frac{y-1}{2\epsilon} \right) \right\}. \end{aligned} \quad (12)$$

In the full-erasure limit ($\epsilon \rightarrow 0$), we obtain the compact expression

$$\Delta_i S_{\min, \text{leq}} = \frac{2}{\tau} \left[1 - \int_0^1 dy f_0^{-1}(y) f_0^{-1} \left(\frac{y}{2} \right) \right]. \quad (13)$$

A. Bounds on optimal local-equilibrium erasure cost

In Appendix B, we show that the remaining integrals in Eq. (12) can be further evaluated, yielding an upper bound on the optimal erasure cost that depends only on the variance of the initial distribution:

$$\Delta_i S_{\min, \text{leq}} \leq 2(1 - 2\epsilon) \frac{\text{Var}(x)}{\tau}. \quad (14)$$

Moreover, Dechant and Sakurai [24] recently showed that the entropy production incurred when transforming a system from an initial state $p_0(x)$ to a final state $p_\tau(x)$ over a period τ is bounded by a quantity that only depends on the first two

moments of the initial and final probability distributions:

$$\tau \Delta_i S_{\min, \text{leq}} \geq \langle x^2 \rangle_0 + \langle x^2 \rangle_\tau - 2\langle x \rangle_0 \langle x \rangle_\tau - 2\sqrt{(\langle x^2 \rangle_0 - \langle x \rangle_0^2)(\langle x^2 \rangle_\tau - \langle x \rangle_\tau^2)}. \quad (15)$$

Here the subscripts 0 and τ indicate averages over $p_0(x)$ and $p_\tau(x)$, respectively. In the context of bit erasure, we can simplify this expression. First, because of the symmetry of $V_0(x)$, $\langle x \rangle_0 = 0$. Second, local equilibrium implies that $\langle x^2 \rangle_\tau = \langle x^2 \rangle_0 = \text{Var}(x)$ [cf. Eq. (A6)]. Finally, from Eq. (8), $\langle x \rangle_\tau = (2\epsilon - 1)\langle |x| \rangle_0$. Equation (15) then simplifies to

$$\Delta_i S_{\min, \text{leq}} \geq \frac{2\text{Var}(x)}{\tau} \left[1 - \sqrt{1 - \frac{(1 - 2\epsilon)^2 \langle |x| \rangle_0^2}{\text{Var}(x)}} \right], \quad (16)$$

which leads to a lower bound for $W_{\min, \text{leq}}$, again in terms of the initial microscopic distribution. Below, we derive a different lower bound that is valid for the work needed in optimal erasure when the final state is *not* constrained to be in local equilibrium.

IV. OPTIMAL ERASURE BEYOND LOCAL EQUILIBRIUM

We now relax the assumption of a local-equilibrium final state and minimize over all final densities $p_\tau(x)$ compatible with the desired macrostate. In contrast to the local-equilibrium protocols from Sec. III, the optimal protocols in this section cannot easily be written in terms of the cumulative distribution functions. Therefore, following Refs. [14, 18–20], we transform the integrals in Eq. (7), which contain quantities evaluated at both the beginning and end of the protocol ($t = 0$ and τ) into quantities evaluated solely at time 0. To change variables carefully, we first refine our notation so that x_0 is the x coordinate used to describe quantities at time $t = 0$, while x_τ is used for quantities at time τ . Further, we define a mapping between them [see Eq. (4)],

$$x_\tau = \Gamma_\tau(x_0) \equiv \Gamma(x_0), \quad (17)$$

so that the cumulative distributions obey $f_\tau(x_\tau) = f_0(x_0)$. Differentiation gives

$$p_\tau(x_\tau) dx_\tau = p_0(x_0) dx_0, \quad p_\tau(x_\tau) = \frac{p_0(x_0)}{\left| \frac{dx_\tau}{dx_0} \right|} = \frac{p_0(x_0)}{\Gamma'(x_0)}. \quad (18)$$

Evaluating the ratio $p_\tau(x_\tau)/p_0(x_\tau)$ that appears in Eq. (7) in terms of x_0 gives

$$\frac{p_\tau(x_\tau)}{p_0(x_\tau)} = \frac{p_0(x_0)}{\Gamma'(x_0)p_0(\Gamma(x_0))}, \quad (19)$$

so that

$$\int_{-\infty}^{\infty} dx_\tau p_\tau(x_\tau) \ln \frac{p_\tau(x_\tau)}{p_0(x_\tau)} = \int_{-\infty}^{\infty} dx_0 p_0(x_0) \ln \left[\frac{p_0(x_0)}{\Gamma'(x_0)p_0(\Gamma(x_0))} \right]. \quad (20)$$

With $y = f_0(x_0) = f_\tau(\Gamma(x_0))$ and $dy = p_0(x_0)dx_0$, the second integral in Eq. (7) becomes

$$\tau \Delta_i S = \int_0^1 dy [f_0^{-1}(y) - f_\tau^{-1}(y)]^2 = \int_{-\infty}^{\infty} dx_0 p_0(x_0) [x_0 - \Gamma(x_0)]^2. \quad (21)$$

To optimize the entropy production with respect to the final distribution, we substitute Eqs. (20) and (21) into Eq. (7). Simplifying the notation from $x_0 \rightarrow x$, we have

$$W_{\min} = \min_{\Gamma(x)} \int_{-\infty}^{\infty} dx p_0(x) \times \left\{ \ln \frac{p_0(x)}{\Gamma'(x)p_0(\Gamma(x))} + \frac{[\Gamma(x) - x]^2}{\tau} \right\}, \quad (22)$$

where W_{\min} is the minimum average work required when $p_\tau(x)$ may vary. In particular, we specify a class of allowed final distributions where the bit is erased with probability $1 - \epsilon$,

$$f_\tau(0) = 1 - \epsilon, \quad (23)$$

In terms of $\Gamma(x)$ this condition can be written as $\Gamma(f_0^{-1}(1 - \epsilon)) = 0$, or

$$\int_{-\infty}^{\infty} dx \Gamma(x) \delta[x - f_0^{-1}(1 - \epsilon)] = 0. \quad (24)$$

From Appendix C, this minimization leads to

$$p_0(\Gamma(x)) \frac{d}{dx} \left[\frac{p_0(x)}{\Gamma'(x)p_0(\Gamma(x))} \right] + \frac{2p_0(x)}{\tau} [\Gamma(x) - x] = \tilde{\lambda} \delta[x - f_0^{-1}(1 - \epsilon)]. \quad (25)$$

If $p_0(x)$ is the Boltzmann distribution corresponding to the potential $V_0(x)$, then we can rewrite Eq. (25) as

$$V_0'(\Gamma(x)) - \frac{V_0'(x)}{\Gamma'(x)} - \frac{\Gamma''(x)}{\Gamma'(x)^2} + \frac{2}{\tau} [\Gamma(x) - x] = \lambda \delta[x - f_0^{-1}(1 - \epsilon)]. \quad (26)$$

Here λ is a parameter that is fixed by the boundary conditions, Eq. (24), and $\Gamma(x) = x$, when $x = \pm\infty$.

A. Lower bound on optimal erasure cost

Minimizing the work, Eq. (7), over all possible final distributions $p_\tau(x)$ requires a trade-off between minimizing the free-energy difference (the Kullback-Leibler distance between initial and final distributions) and the entropy production. Minimizing the former leads to the local-equilibrium distribution in Eq. (8), which in turn implies the free-energy difference given in Eq. (10).

Borrowing from Eq. (22), the latter corresponds to minimizing

$$\int_{-\infty}^{\infty} dx p_0(x) \frac{[\Gamma(x) - x]^2}{\tau} \quad (27)$$

with respect to $\Gamma(x)$, under the constraints that $\Gamma(x)$ is nondecreasing and that Eq. (24) holds. The constraints correspond to setting $\Gamma(x) \leq 0$ for $x < f_0^{-1}(1 - \epsilon)$ and $\Gamma(x) \geq 0$ for

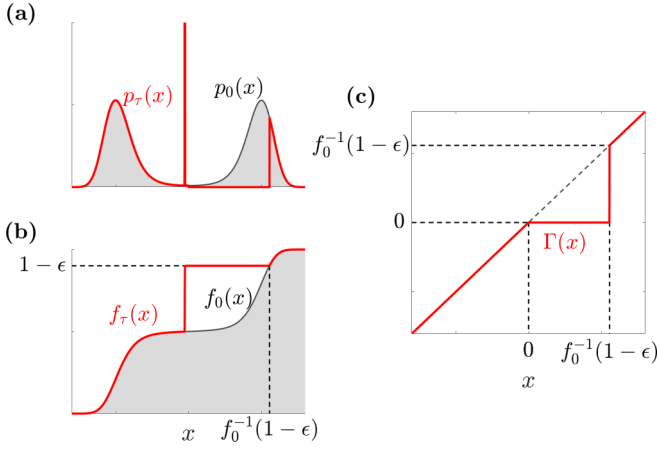


FIG. 1. Illustration of the final distribution compatible with the final macrostate that leads to minimum entropy production. (a) Probability density functions $p_0(x)$ (red) and $p_\tau(x)$ (dark gray). (b) Cumulative distributions $f_0(x)$ (red) and $f_\tau(x)$ (dark gray). (c) The function $\Gamma(x)$ (red) shows the coordinate transformation from time 0 to τ . The curves are calculated for a quartic well [see Eq. (37)] with $E_b = 4$ and $\epsilon = 0.1$.

$x > f_0^{-1}(1 - \epsilon)$. Minimizing Eq. (27) then corresponds to minimizing the distance between x and $\Gamma(x)$. As a result, the optimal $\Gamma(x)$ is

$$\Gamma(x) = \begin{cases} x, & x < 0 \\ 0^-, & 0 \leq x \leq f_0^{-1}(1 - \epsilon), \\ x, & x > f_0^{-1}(1 - \epsilon) \end{cases} \quad (28)$$

as shown in Fig. 1. We can write this minimization as

$$\begin{aligned} \min_{\Gamma(x)} \int_{-\infty}^{\infty} dx p_0(x) \frac{[\Gamma(x) - x]^2}{\tau} \\ = \frac{1}{\tau} \int_0^{f_0^{-1}(1-\epsilon)} dx p_0(x) x^2. \end{aligned} \quad (29)$$

This amounts to “pushing” the necessary amount of probability density “just over” the barrier separating the two macrostates while “leaving behind” the probability ϵ remaining after partial erasure (Fig. 1).

A lower bound for the minimum work needed is then given by combining the individual minimizations:

$$\begin{aligned} W_{\min} &\geq \min_{p_\tau(x)} \int_{-\infty}^{\infty} dx p_\tau(x) \ln \frac{p_\tau(x)}{p_0(x)} \\ &\quad + \min_{\Gamma(x)} \int_{-\infty}^{\infty} dx p_0(x) \frac{[\Gamma(x) - x]^2}{\tau} \\ &= \ln 2 + \epsilon \ln \epsilon + (1 - \epsilon) \ln(1 - \epsilon) \\ &\quad + \frac{1}{\tau} \int_0^{f_0^{-1}(1-\epsilon)} dx p_0(x) x^2. \end{aligned} \quad (30)$$

This bound is generally saturated in the limit $\tau \rightarrow 0$ because of the overwhelming contribution of the entropy production to the total work.

In the limit $\epsilon \rightarrow 0$ of full erasure, we find

$$W_{\min} \geq \ln 2 + \frac{\text{Var}(x)}{2\tau}, \quad (31)$$

as discussed in the accompanying paper [16].

V. EXAMPLES

In this section, we apply our general analysis to two example systems, with *flat-well* and *quartic* potentials.

A. Flat well

Let us consider information erasure in a system consisting of a particle initially in equilibrium with a potential consisting of two flat wells, separated by a thin barrier. This particular example has been studied in Ref. [12] under limited control and has the advantage that it can be evaluated analytically. For the width of the individual wells, we choose $\sqrt{3}$, so that $\langle x^2 \rangle_0 = 1$.

In Appendix D, we show that the erasure protocol that leads to a local-equilibrium final distribution has the intermediate probability distribution,

$$p(x, t) = \begin{cases} \frac{\sqrt{3}(1-\epsilon)}{6(1-\epsilon)-3(1-2\epsilon)t/\tau} & x < x_0(t) \\ \frac{\sqrt{3}\epsilon}{3(1-2\epsilon)t/\tau + 6\epsilon} & x > x_0(t) \end{cases}, \quad (32)$$

where

$$x_0(t) = \sqrt{3}(1 - 2\epsilon)(1 - t/\tau). \quad (33)$$

Remarkably, the optimum intermediate probability distribution stays piecewise uniform throughout the erasure process.

We also calculate the intermediate potential by inverting the Fokker-Planck equation, Eq. (6), which leads to Eq. (D8). The potential is illustrated in Fig. 2(a) along with the intermediate probability distribution.

As detailed in Appendix D, the entropy production when erasing to local equilibrium is given by

$$\Delta_i S_{\min, \text{leq}} = \frac{(1 - 2\epsilon)^2}{\tau}. \quad (34)$$

The quadratic dependence of the entropy production on the erasure error differs from the expressions found when the form of the potential $V(x, t)$ is constrained in parametric optimization [12] and when the density is constrained to be at local equilibrium for $0 \leq t \leq \tau$ in counterdiabatic driving [15]. In our case, local equilibrium is imposed only at $t = 0$ and $t = \tau$.

Equations (14), (16), and (34) then become the following set of inequalities:

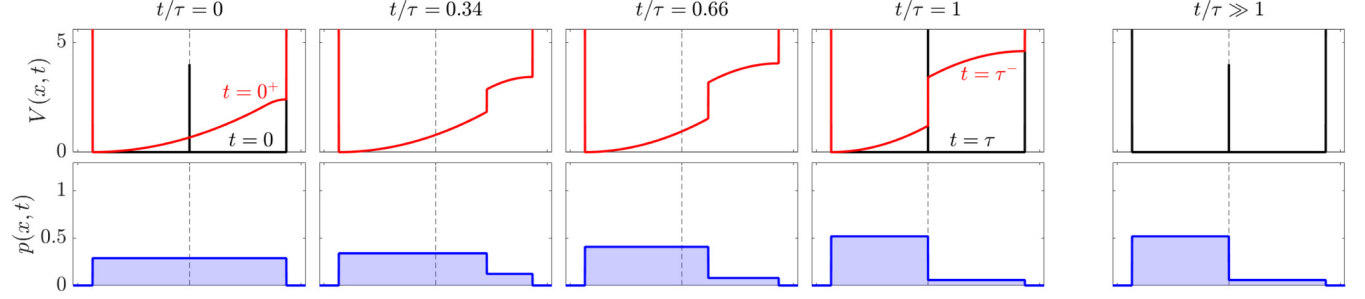
$$\begin{aligned} 2 - 2\sqrt{1 - \frac{3(1-2\epsilon)^2}{4}} &\leq \Delta_i S_{\min, \text{leq}} \tau = (1 - 2\epsilon)^2 \\ &\leq 2(1 - 2\epsilon). \end{aligned} \quad (35)$$

Figure 3 shows the entropy production together with the upper and lower bound. We immediately see that both bounds are verified and that the lower bound becomes tight in the full-erasure and zero-erasure limits.

1. Erasure beyond local equilibrium

Although a flat-well potential is arguably the simplest case for bit erasure, it is still not possible to solve Eq. (26) analytically. Therefore, we turn to the numerical methods developed in Appendix E. The resulting optimal erasure process

(a) Erasure to local equilibrium:



(b) Optimal erasure:

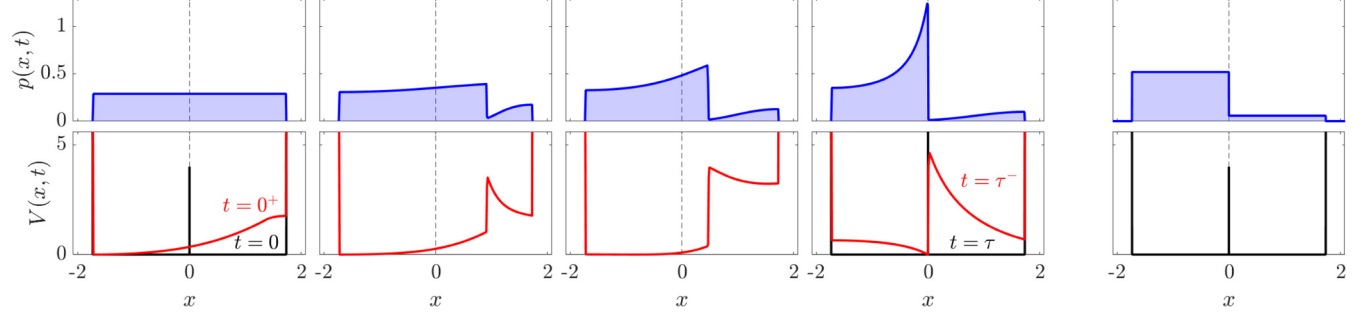


FIG. 2. Partial ($\epsilon = 0.1$) finite-time ($\tau = 1$) erasure of a memory consisting of two flat wells separated by a thin barrier. (a) Potential $V(x, t)$ achieving a local-equilibrium final distribution. Second row: Resulting intermediate probability distribution $p(x, t)$. Third row: Intermediate probability distribution for optimal erasure. (b) Potential achieving optimum erasure. At $t = \tau$ we prevent immediate relaxation to global equilibrium by including a potential barrier at $x = 0$.

is illustrated in Fig. 2(b). In this protocol, a barrier emerges in the right well during the first part of the protocol. This barrier then steadily moves to the center of the potential, pushing the probability to the left. As τ is small, the system does not have time to relax to local equilibrium, leading to a peak in the probability distribution near the barrier. After completing the protocol, the system relaxes back to the local-equilibrium distribution for $t \gg \tau$.

The entropy production for these various erasure protocols is shown in Fig. 4. For slow driving $W_{\min} \approx W_{\min, \text{leq}}$ (dashed

red line). Meanwhile, for fast driving, $\tau \rightarrow 0$, W_{\min} converges to the lower bound [lower black solid line, given by Eq. (30)].

For a flat-well potential, this lower bound can be evaluated analytically using Eq. (D9),

$$\Delta_i S_{\min} \geq \frac{(1 - 2\epsilon)^3}{2}. \quad (36)$$

Comparing this result with Eq. (34) shows that the lower bound differs from the local-equilibrium result by a factor $2/(1 - 2\epsilon)$. This means that relaxing the assumption of local equilibrium is particularly useful in the fast driving limit.

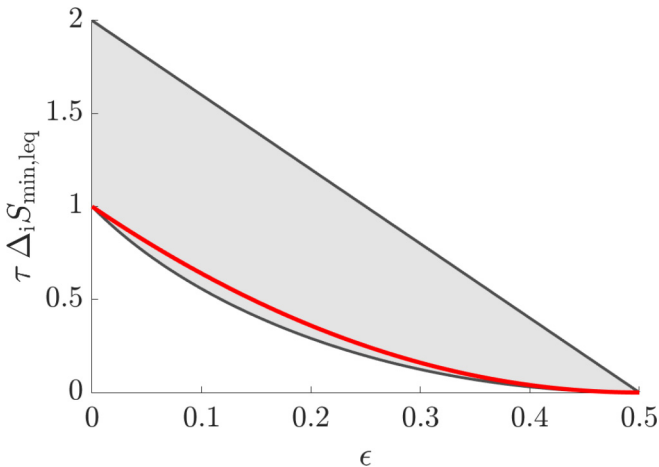


FIG. 3. Comparison between entropy production (thick red curve) and its upper and lower bounds (gray curves bounding the shaded region) for two flat wells, separated by a thin barrier.

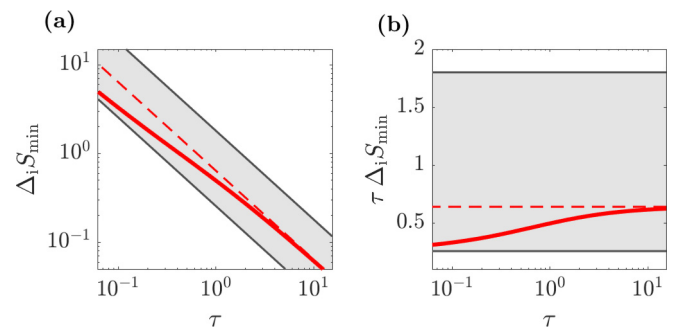


FIG. 4. Finite-time erasure costs in a flat-well potential. (a) Comparison between $\Delta_i S_{\min}$ (full line), $\Delta_i S_{\min, \text{leq}}$ (dashed line) for partial erasure ($\epsilon = 0.1$) of a flat well. Additionally, the upper and lower bounds, Eqs. (14) and (30) (shaded region) are shown. (b) Same quantities as the left panel, but scaled by τ .

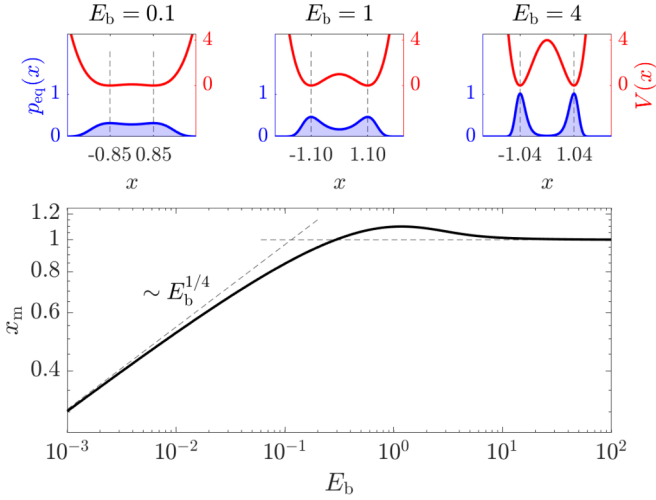


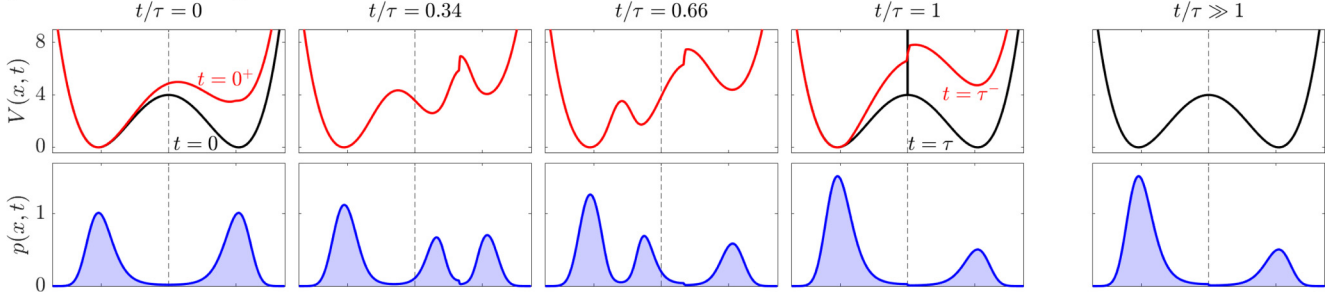
FIG. 5. Double-well potentials with different barrier heights. (a) Equilibrium probability densities (blue shaded) for double-well potentials of energy barriers $E_b = 0.1, 1, 3$, in units of $k_B T$ (red curves). (b) Plot of the value of x_m that makes $\text{Var}_0(x) = 1$, as a function of E_b . Dotted line indicates that $x_m \sim (E_b)^{1/4}$ as $E_b \rightarrow 0$.

B. Quartic double well

Next consider the potential

$$V_0(x) = E_b \left[\left(\frac{x}{x_m} \right)^2 - 1 \right]^2, \quad (37)$$

(a) Erasure to local equilibrium:



(b) Optimal erasure:

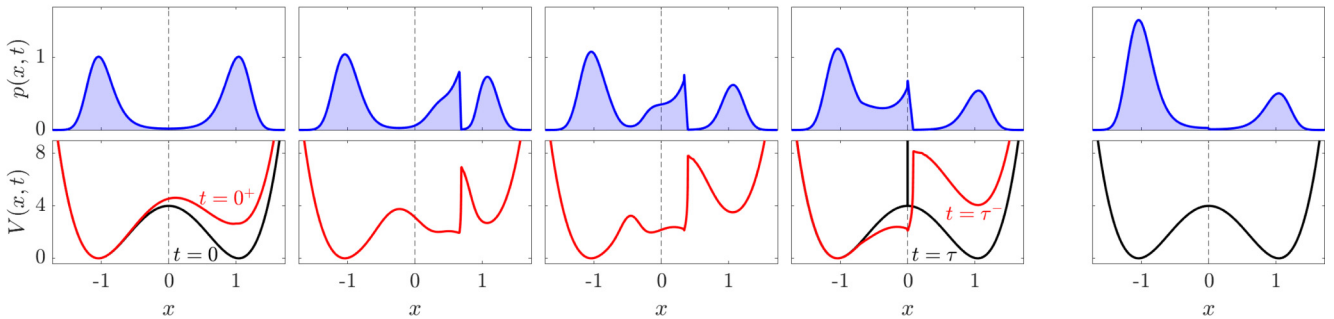


FIG. 6. Partial ($\epsilon = 0.25$) finite-time ($\tau = 0.5$) erasure of a memory consisting of a quartic double well with energy barrier $E_b = 4$. (a) Potential $V(x, t)$ achieving a local-equilibrium final distribution and resulting intermediate probability distribution $p(x, t)$. (b) Intermediate probability distribution for optimal erasure and potential achieving optimum erasure. At $t = \tau$ we prevent immediate relaxation to global equilibrium by including a potential barrier at $x = 0$. At $t/\tau = 1$, the jump in the potential is slightly shifted to positive values due to numeric instabilities.

which corresponds to a double-well potential, with energy barrier E_b and minima at $x = \pm x_m$ (see Fig. 5). This potential has been studied extensively in the context of bit erasure, both experimentally [6–9,11] and numerically [14,15].

Here we choose x_m as a function of E_b so that $\text{Var}(x) = 1$ for all barrier heights. The function $x_m(E_b)$ is plotted in Fig. 5 (black curve). For $E_b \ll 1$, a scaling analysis of the condition $\langle x^2 \rangle = 1$ shows that $x_m \approx [\Gamma(1/4)/(2^{1/4}\pi^{1/2})](E_b)^{1/4} \approx 1.72(E_b)^{1/4}$. Conversely, for $E_b \gg 1$, the equilibrium distribution tends to two δ functions at $\pm x_m$, implying that $x_m \rightarrow 1$.

The erasure protocol $V(x, t)$ leading to a local-equilibrium final distribution for $E_b = 4$ can only be calculated numerically (see Appendix E). It is illustrated in Fig. 6(a).

We numerically calculate the dissipative correction $\Delta_i S_{\text{min,leq}}$ to the bare Landauer cost from Eq. (12). Additionally, using Eqs. (14) and (16), we numerically calculate the upper and lower bounds for the entropy production.

The results are displayed in Fig. 7. Importantly, the additional cost for optimal finite-time erasure does not scale exponentially with barrier height as claimed in Ref. [15]. Instead, we see that increasing the barrier height leads to a saturation of the upper bound. This limiting case corresponds to a potential with very deep wells, which leads to an initial distribution consisting of two “delta peaks” [16].

1. Erasure beyond local equilibrium

Figure 6(b) illustrates the optimum erasure process. Then, in Fig. 8, we show the (numerically calculated) optimal work for finite-time erasure. One can check that, for fixed variance of the initial distribution, the erasure cost grows with a higher

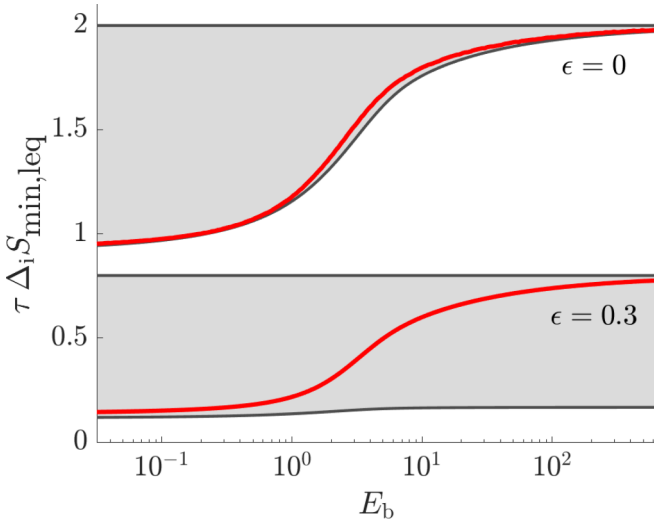


FIG. 7. Entropy production (solid lines) for partial erasure of the quartic double-well potential together with the upper and lower bounds (shaded region) as a function of the barrier height for full and partial erasure.

barrier, both in the fast and in the slow erasure limit. In the slow limit, we again find that W_{\min} approaches $W_{\min, \text{leq}}$. In the fast limit, it approaches the lower bound.

VI. DISCUSSION AND CONCLUSION

In this article, we have given a detailed treatment of the theory underlying optimization of thermodynamic processes, where the final constraint is placed at the level of a macrostate. At the level of microstates, an entire class of distributions is allowed. We have applied this formalism to the case of information erasure, deriving both optimal protocols and various bounds.

Optimizing over a class of microscopic distributions can yield an optimal protocol for finite time bit erasure that uses up to a factor of four less work compared to protocols that end in local-equilibrium when the bit is fully erased [16]. When only part of a bit is erased, the difference in erasure cost between the local equilibrium protocol and the optimal one can be even greater, a conclusion that follows from our analysis of the flat-well potential: For short erasure processes, the minimum erasure cost tends toward the lower bound in Eq. (36), while the local-equilibrium erasure cost is given by Eq. (34). Their ratio is a factor of $2/(1 - 2\epsilon)$, which diverges for $\epsilon \rightarrow 1/2$.

The method for finding the final probability distribution that results in the minimum-work erasure protocol is quite involved and can only be treated numerically even for quite simple example systems. Nonetheless, the resulting time-evolution of the probability distribution has some intuitive appeal (see Figs. 2 and 6): An optimal partial erasure protocol into the left well takes the closest part of the probability distribution from the right well, pushes it over the barrier and redistributes probability density so as to minimize a trade-off between free-energy difference and entropy production.

In that context, it is noteworthy that the control potential $V(x, t)$ needed to achieve a desired evolution $p(x, t)$ of prob-

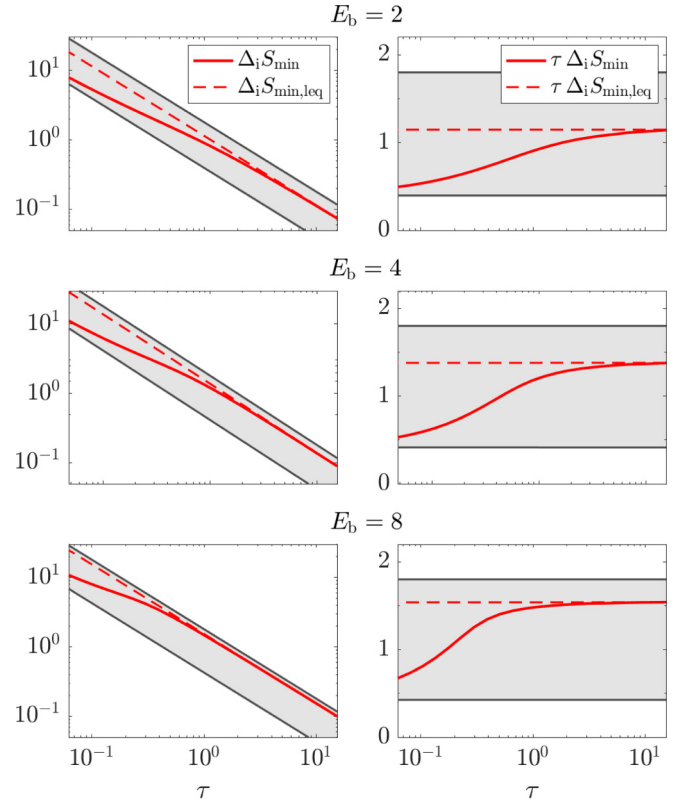


FIG. 8. Finite-time erasure costs in a quartic potential. (a) Comparison between $\Delta_i S_{\min}$ (full line), $\Delta_i S_{\min, \text{leq}}$ (dashed line) and the upper and lower bounds, Eqs. (14) and (30) (shaded region) for a double well, Eq. (37) with $E_b = 2, 4, 8$. (b) Same quantities as the left panel but scaled by τ .

ability density is often quite involved, with steep gradients and discontinuities. The complicated shape of the required potential leads not only to numerical difficulties (Appendix E) but also to experimental challenges. Indeed, an important open question is whether inevitable limitations to experimental control over the potential can nonetheless allow experimental protocols whose average work closely approximate the minimum values discussed here.

One straightforward strategy to further reduce costs beyond our derived bounds would be to manipulate the intrinsic time and length scales during the protocol. In the limit of infinitely fast relaxation times, such a strategy would enable finite-time erasure at no additional cost beyond the Landauer limit as shown, e.g., in Refs. [25,26]. Tempting as such an approach might be, it is not realistic. Although reducing the length and time scales improves the performance of bit erasure, it would make little sense to have the ability to impose faster or shorter time scales and only use them for part of a protocol. Thus, we consider such a *technological* optimization to have been first carried out, fixing length and time scales. Then one proceeds to *protocol* optimization, the focus of this paper.

One potential drawback of our finite-time protocols is that, for fast erasure, probability accumulates near the center of the potential and might leak back into the wrong part of the potential after completion of the protocol. There are several ways to avoid this problem: As already mentioned in Sec. V,

one could add a thin, high barrier at the center of the potential immediately after completing the erasure. The barrier would guarantee that the system relaxes to the correct macrostate and can be removed later. Another possibility would be to push the probability slightly farther, past the center of the potential. In this way, the system will generally relax to the correct state, with a controllable trade-off between the extra work required to push probability farther past the edge of the macrostate, against the amount of probability that will leak back into the original macrostate. Similarly, if the goal is partial erasure, then one can choose a protocol with a nominally smaller value of ϵ that will reach the desired value in the relaxation after the end of the protocol. In this article, we have not explored such trade-offs, which all constitute suboptimal protocols. Our reasoning is that there are many ways to be suboptimal, and it makes more sense to explore them in the context of a specific experimental realization, where experimental constraints, which also lead to suboptimal strategies, can be accommodated. In the absence of specific experimental constraints, our strategy, which achieves minimum work at the cost of needing extra time afterward, seems a reasonable exemplar.

Finally, we note that our study here and in Ref. [16] has focused on the average work to perform the single operation of bit erasure. It would be interesting to consider the trade-offs between mean values and fluctuations [27,28]. The main difficulty in optimizing a trade-off between the average amount of work and its variance is that one no longer has a direct connection with the Wasserstein distance. This makes the application of optimal transport theory significantly more challenging. Furthermore, it would be interesting to look for optimizations that include a sequence of computational operations [29]. To do this, one would need to take into account that at the start of a new operation, the bit might not yet be in local equilibrium. Therefore, to optimize the full sequence of operations, every separate operation might have to be suboptimal.

ACKNOWLEDGMENTS

This work was supported by Foundational Questions Institute Grant No. FQXi-RFP-2019-IAF and by a National Science and Engineering Research Council of Canada Discovery Grant.

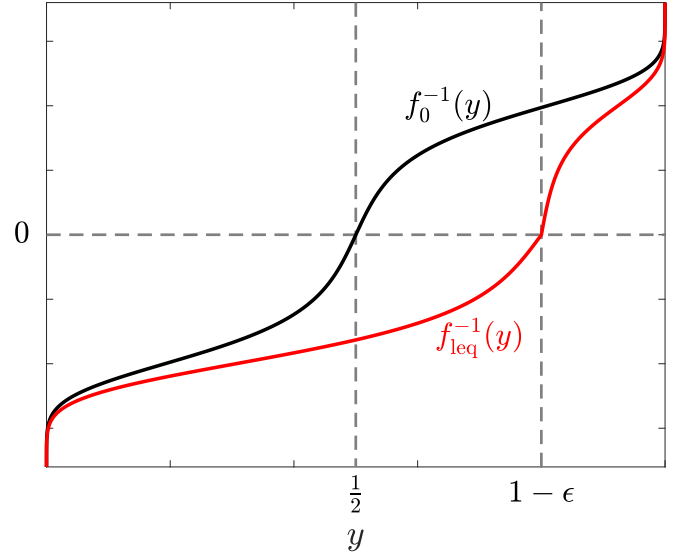


FIG. 9. Example plot of the initial inverse distribution function $f_0^{-1}(y)$ and the final local-equilibrium inverse distribution function $f_{leq}^{-1}(y)$. The curves are calculated for a quartic well [see Eq. (37)] with $E_b = 2$ and $\epsilon = 0.2$.

APPENDIX A: DERIVATION OF EQ. (12)

According to Eq. (8), the cumulative local-equilibrium distribution $f_{leq}(x)$ is given by

$$f_{leq}(x) = \begin{cases} 2(1-\epsilon)f_0(x) & x < 0 \\ 1-\epsilon + 2\epsilon \int_0^x dx' p_0(x') & x \geq 0 \end{cases}, \quad (\text{A1})$$

$$= \begin{cases} 2(1-\epsilon)f_0(x) & x < 0 \\ 1 + 2\epsilon [f_0(x) - 1] & x \geq 0 \end{cases}. \quad (\text{A2})$$

Inverting this function gives

$$f_{leq}^{-1}(y) = \begin{cases} f_0^{-1}\left[\frac{y}{2(1-\epsilon)}\right] & 0 \leq y < 1-\epsilon \\ f_0^{-1}\left(1 + \frac{y-1}{2\epsilon}\right) & 1-\epsilon \leq y \leq 1 \end{cases}. \quad (\text{A3})$$

Figure 9 shows an example plot of the resulting inverse functions.

Expanding the integrand in Eq. (11) and inserting Eq. (A3) gives

$$\Delta_i S_{\min,leq} = \frac{1}{\tau} \left\{ \int_0^1 dy f_0^{-1}(y)^2 + \int_0^{1-\epsilon} dy f_0^{-1}\left[\frac{y}{2(1-\epsilon)}\right]^2 + \int_{1-\epsilon}^1 dy f_0^{-1}\left(1 + \frac{y-1}{2\epsilon}\right)^2 - 2 \int_0^1 dy f_0^{-1}(y) f_{leq}^{-1}(y) \right\}. \quad (\text{A4})$$

Substituting $y = f_0(x)$ in the first integral, $y/2(1-\epsilon) = f_0(x)$ in the second integral, and $1 + (y-1)/2\epsilon = f_0(x)$ in the third integral yields

$$\begin{aligned} \Delta_i S_{\min,leq} &= \frac{1}{\tau} \left[\int_{-\infty}^{\infty} dx p_0(x) x^2 + 2(1-\epsilon) \int_{-\infty}^0 dx p_0(x) x^2 + 2\epsilon \int_0^{\infty} dx p_0(x) x^2 - 2 \int_0^1 dy f_0^{-1}(y) f_{leq}^{-1}(y) \right] \\ &= \frac{2}{\tau} \left[\langle x^2 \rangle_0 - \int_0^1 dy f_0^{-1}(y) f_{leq}^{-1}(y) \right], \end{aligned} \quad (\text{A5})$$

where $\langle x^2 \rangle_0$ is the second moment of the initial distribution. Because of the symmetry of the potential (and therefore the initial distribution), it equals the variance of the initial distribution:

$$\langle x^2 \rangle_0 = \text{Var}(x) \equiv 1. \quad (\text{A6})$$

Finally, inserting Eq. (A3) for the remaining integral and recalling Eqs. (9) and (10), we obtain Eq. (12) of the main text.

APPENDIX B: DERIVATION OF EQ. (14)

We start by splitting the integrals in Eq. (12) further to obtain

$$\int_0^{1-\epsilon} dy f_0^{-1}(y) f_0^{-1} \left[\frac{y}{2(1-\epsilon)} \right] = I_1 + I_2 \quad (\text{B1})$$

and

$$\int_{1-\epsilon}^1 dy f_0^{-1}(y) f_0^{-1} \left(1 + \frac{y-1}{2\epsilon} \right) = I_3 + I_4, \quad (\text{B2})$$

where

$$I_1 \equiv \int_0^\epsilon dy f_0^{-1}(y) f_0^{-1} \left[\frac{y}{2(1-\epsilon)} \right] \geq \int_0^\epsilon dy f_0^{-1}(y)^2, \quad (\text{B3})$$

$$I_2 \equiv \int_\epsilon^{1/2} dy f_0^{-1}(y) f_0^{-1} \left[\frac{y}{2(1-\epsilon)} \right], \quad (\text{B4})$$

and

$$I_3 \equiv \int_{1/2}^{1-\epsilon} dy f_0^{-1}(y) f_0^{-1} \left[\frac{y}{2(1-\epsilon)} \right], \quad (\text{B5})$$

$$I_4 \equiv \int_{1-\epsilon}^1 dy f_0^{-1}(y) f_0^{-1} \left(1 + \frac{y-1}{2\epsilon} \right) \geq \int_{1-\epsilon}^1 dy f_0^{-1} \left(1 + \frac{y-1}{2\epsilon} \right)^2. \quad (\text{B6})$$

The inequalities follow from the fact that $f_0^{-1}(y)$ is an increasing function and $\epsilon \leq 1/2$ (see also Fig. 9).

Because $p_0(x)$ is an even function, we have $p_0(-x) = p_0(x)$, and the cumulative distribution obeys

$$f_0(-x) + f_0(x) = 1, \quad (\text{B7})$$

which in turn implies

$$\begin{aligned} f_0(-f_0^{-1}(y)) &= 1 - f_0(f_0^{-1}(y)) \\ \iff f_0^{-1}(y) &= -f_0^{-1}(1-y). \end{aligned} \quad (\text{B8})$$

Equation (B8) can be used to rewrite I_2 :

$$I_2 = - \int_\epsilon^{1/2} dy f_0^{-1}(1-y) f_0^{-1} \left[\frac{y}{2(1-\epsilon)} \right] = - \int_{1/2}^{1-\epsilon} dy' f_0^{-1}(y') f_0^{-1} \left[\frac{1-y'}{2(1-\epsilon)} \right], \quad (\text{B9})$$

which implies

$$I_2 + I_3 = \int_{\frac{1}{2}}^{1-\epsilon} dy f_0^{-1}(y) \left\{ f_0^{-1} \left[\frac{y}{2(1-\epsilon)} \right] - f_0^{-1} \left[\frac{1-y}{2(1-\epsilon)} \right] \right\}, \quad (\text{B10})$$

Note that $f_0^{-1}(y) > 0$ for $y > 1/2$ and recall that $f_0^{-1}(y)$ is an increasing function. Since $y \geq 1-y$ in the interval $[1/2, 1]$, it is clear that the integrand in Eq. (B10) is positive over the entire integration domain. From this, we can deduce

$$I_2 + I_3 \geq 0. \quad (\text{B11})$$

Next, combining I_1 and I_4 yields

$$\begin{aligned} I_1 + I_4 &\geq \int_0^\epsilon dy f_0^{-1}(y)^2 + \int_{1-\epsilon}^1 dy f_0^{-1} \left(1 + \frac{y-1}{2\epsilon} \right)^2 = \int_{1-\epsilon}^1 dy' f_0^{-1}(y')^2 + \int_{1-\epsilon}^1 dy f_0^{-1} \left(1 + \frac{y-1}{2\epsilon} \right)^2 \\ &\geq 2 \int_{1-\epsilon}^1 dy f_0^{-1} \left(1 + \frac{y-1}{2\epsilon} \right)^2 = 4\epsilon \int_{1/2}^1 dy' f_0^{-1}(y')^2 = 2\epsilon \langle x^2 \rangle_0, \end{aligned} \quad (\text{B12})$$

where we used the symmetry relation, Eq. (B8), in line 2 and the fact that $f_0^{-1}[1 + (y-1)/2\epsilon]$ is smaller than $f_0^{-1}(y)$ (see Fig. 9) in line 3.

Equations (12), (A6), (B1), (B2), (B11), and (B12) together lead to

$$\Delta_i S_{\min, \text{leq}} \leq 2(1-2\epsilon) \frac{\text{Var}(x)}{\tau}, \quad (\text{B13})$$

which is Eq. (14) in the main text.

APPENDIX C: DERIVATION OF EQ. (25)

To minimize Eq. (22) under the constraint Eq. (24), we define an augmented Lagrangian

$$\mathcal{L} = \int_{-\infty}^{\infty} dx p_0(x) \left[\ln \frac{p_0(x)}{\Gamma'(x)p_0(\Gamma(x))} + \frac{[\Gamma(x) - x]^2}{\tau} \right] - \tilde{\lambda} \int_{-\infty}^{\infty} dx \Gamma(x) \delta[x - f_0^{-1}(1 - \epsilon)], \quad (\text{C1})$$

and set variations with respect to $\Gamma(x)$ equal to zero,

$$\frac{\delta \mathcal{L}}{\delta \Gamma(x)} = 0. \quad (\text{C2})$$

One can verify that, to first order in $\delta \Gamma(x)$,

$$\begin{aligned} & \int_{-\infty}^{\infty} dx p_0(x) \ln \frac{p_0(x)}{[\Gamma'(x) + \delta \Gamma'(x)]p_0[\Gamma(x) + \delta \Gamma(x)]} \\ &= \int_{-\infty}^{\infty} dx p_0(x) \ln \frac{p_0(x)}{\Gamma'(x)p_0(\Gamma(x))} - \int_{-\infty}^{\infty} dx \frac{p_0(x)}{\Gamma'(x)} \delta \Gamma'(x) - \int_{-\infty}^{\infty} dx \frac{p_0(x)p_0'(\Gamma(x))}{p_0(\Gamma(x))} \delta \Gamma(x) \\ &= \int_{-\infty}^{\infty} dx p_0(x) \ln \frac{p_0(x)}{\Gamma'(x)p_0(\Gamma(x))} + \int_{-\infty}^{\infty} dx \delta \Gamma(x) p_0(\Gamma(x)) \frac{d}{dx} \frac{p_0(x)}{\Gamma'(x)p_0(\Gamma(x))}, \end{aligned} \quad (\text{C3})$$

where we integrated by parts in the second integral of the second line to arrive at the last line. Furthermore, again to first order in $\delta \Gamma(x)$,

$$\int_{-\infty}^{\infty} dx p_0(x) [\Gamma(x) + \delta \Gamma(x) - x]^2 = \int_{-\infty}^{\infty} dx p_0(x) [\Gamma(x) - x]^2 + 2 \int_{-\infty}^{\infty} dx p_0(x) [\Gamma(x) - x] \delta \Gamma(x)$$

and

$$\int_{-\infty}^{\infty} dx [\Gamma(x) + \delta \Gamma(x)] \delta[x - f_0^{-1}(1 - \epsilon)] = \int_{-\infty}^{\infty} dx \Gamma(x) \delta[x - f_0^{-1}(1 - \epsilon)] + \int_{-\infty}^{\infty} dx \delta \Gamma(x) \delta[x - f_0^{-1}(1 - \epsilon)], \quad (\text{C4})$$

which leads to a variation

$$\delta \mathcal{L} = \int_{-\infty}^{\infty} dx \delta \Gamma(x) \left\{ p_0(\Gamma(x)) \frac{d}{dx} \frac{p_0(x)}{\Gamma'(x)p_0(\Gamma(x))} + \frac{2p_0(x)}{\tau} [\Gamma(x) - x] - \tilde{\lambda} \delta[x - f_0^{-1}(1 - \epsilon)] \right\}. \quad (\text{C5})$$

The Euler-Lagrange equation, Eq. (C2), now leads to Eq. (25) in the main text.

APPENDIX D: EXACT CALCULATIONS FOR THE FLAT-WELL POTENTIAL

Here we present the detailed calculations for the results presented in Sec. V A.

1. Intermediate probability distribution and potential

The initial cumulative distribution is given by

$$f_0(x) = \frac{x + \sqrt{3}}{2\sqrt{3}} \quad \text{for } |x| \leq \sqrt{3}. \quad (\text{D1})$$

With this and Eq. (A2), we eventually find the inverse of the interpolating function, Eq. (4),

$$\Gamma_t^{-1}(x) = \begin{cases} \frac{\sqrt{3}(1-2\epsilon)t/\tau + 2(1-\epsilon)x}{2-(1-2\epsilon)t/\tau - 2\epsilon} & x < x_0(t) \\ \frac{\sqrt{3}(1-2\epsilon)t/\tau + 2\epsilon x}{(1-2\epsilon)t/\tau + 2\epsilon} & x > x_0(t) \end{cases}, \quad (\text{D2})$$

where

$$x_0(t) = \sqrt{3}(1 - 2\epsilon)(1 - t/\tau) \quad (\text{D3})$$

is chosen so that $\Gamma_t^{-1}(x)$ and

$$f(x, t) = f_0(\Gamma_t^{-1}(x)) \quad (\text{D4})$$

are continuous.

Differentiating the cumulative distribution yields the intermediate probability distribution in Eq. (32):

$$p(x, t) = \begin{cases} p_l(t) & x < x_0(t) \\ p_r(t) & x > x_0(t) \end{cases} \quad (\text{D5})$$

$$= \begin{cases} \frac{\sqrt{3}(1-\epsilon)}{6(1-\epsilon) - 3(1-2\epsilon)t/\tau} & x < x_0(t) \\ \frac{\sqrt{3}\epsilon}{3(1-2\epsilon)t/\tau + 6\epsilon} & x > x_0(t) \end{cases}. \quad (\text{D6})$$

Using Eq. (6), we calculate the intermediate potential

$$V(x, t) = \begin{cases} \frac{\dot{p}_l(t)}{2p_l(t)}(x + \sqrt{3})^2 & x < x_0(t) \\ V_r(x, t) & x > x_0(t) \end{cases}, \quad (\text{D7})$$

with

$$\begin{aligned} V_r(x, t) \equiv & \ln \frac{p_l(t)}{p_r(t)} + \frac{\dot{p}_l(t)}{2p_l(t)} [x_0(t) + \sqrt{3}]^2 \\ & + \frac{x - x_0(t)}{p_r(t)} \left\{ \dot{p}_l(t) [x_0(t) + \sqrt{3}] \right. \\ & \left. + \frac{\dot{p}_r(t)}{2} [x - x_0(t)] - \dot{x}_0(t) [p_r(t) - p_l(t)] \right\}. \end{aligned} \quad (\text{D8})$$

Note that the remaining constant is chosen such that the minimum of the potential is equal to zero.

2. Entropy production during erasure to local equilibrium

Inverting Eq. (D1) yields

$$f_0^{-1}(y) = \sqrt{3}(2y - 1) \quad \text{for } 0 \leq y \leq 1. \quad (\text{D9})$$

with which Eq. (12) can be evaluated to find Eq. (34) in the main text.

To calculate the lower bound, Eq. (16), we need the following quantity:

$$\langle |x| \rangle_0 = \int_{-\sqrt{3}}^{\sqrt{3}} dx \frac{|x|}{2\sqrt{3}} = \frac{\sqrt{3}}{2}. \quad (\text{D10})$$

APPENDIX E: NUMERICAL METHODS FOR FINDING THE OPTIMAL PROTOCOL AND COMPUTING THE ENTROPY PRODUCTION

To find optimal protocols and associated quantities numerically, we need to solve the variational equation, (26), which defines a boundary-value problem. To simplify the numerics, we limit the domain of the solution by setting $V_0(x) \rightarrow \infty$ for $x \notin [-x_{\max}, x_{\max}]$. Then, we calculate $\Gamma(x)$ via a shooting method [30]: We fix $\Gamma(-x_{\max}) = -x_{\max}$ and make an initial guess for $\Gamma'(-x_{\max})$. We integrate Eq. (26) using a Runge-Kutta method and check whether the condition $\Gamma(f_0^{-1}(1 - \epsilon)) = 0$, cf. Eq. (24), is satisfied. If $\Gamma(\Gamma(f_0^{-1}(1 - \epsilon))) > 0$, we reduce $\Gamma'(-x_{\max})$ and integrate again. Similarly, if $\Gamma(\Gamma(f_0^{-1}(1 - \epsilon))) < 0$ we increase $\Gamma'(-x_{\max})$ and repeat the integration. We iterate until $|\Gamma(x_{\max}) - x_{\max}| < 10^{-3}$. As

Eq. (26) allows for a discontinuity in $\Gamma'(x)$ at $x = f_0^{-1}(1 - \epsilon)$, we can repeat the above procedure to find $\Gamma'(f_0^{-1}(1 - \epsilon))$ that leads to $\Gamma(x_{\max}) = x_{\max}$.

Having determined $\Gamma(x)$, the remaining tasks are to find the time-dependent probability densities $p(x, t)$, the cumulative distribution $f(x, t)$, and the time-dependent potential $V(x, t)$ that implements the control. To find the cumulative distribution $f(x, t)$, we use Eqs. (3) with (4), which translates to

$$\Gamma_t(x) \equiv x + \frac{t}{\tau} [\Gamma(x) - x]. \quad (\text{E1})$$

Having found $f(x, t)$, we can differentiate with respect to x to find the corresponding density $p(x, t)$. To find $V(x, t)$, we use Eq. (6) and change the lower integration limits to $-x_{\max}$. Finally, we compute $\Delta_i S_{\min}$ using Eq. (21). The code (in C) can be found in Ref. [31].

Because the map $\Gamma(x)$ can tend toward functions with discontinuous derivatives—especially for short protocols with $\tau \ll 1$ —finding the numerical solution can be challenging. We checked the shooting method described above with the built-in Mathematica command `NDSolveValue`, which claims to internally implement the shooting method as a root-finding problem.

In addition, we discretized $p_\tau(x)$ and wrote code to estimate W using Eq. (7). We then minimized W by gradient descent, directly varying the discretized values of $p_\tau(x)$. Both of these approaches produced results that agree with those presented in the paper. However, all methods fail for small-enough τ and/or large-enough E_b , where solutions can become more singular.

-
- [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] R. Dillenschneider and E. Lutz, Memory Erasure in Small Systems, *Phys. Rev. Lett.* **102**, 210601 (2009).
 - [4] B. Lambson, D. Carlton, and J. Bokor, Exploring the Thermodynamic Limits of Computation in Integrated Systems: Magnetic Memory, Nanomagnetic Logic, and the Landauer Limit, *Phys. Rev. Lett.* **107**, 010604 (2011).
 - [5] G. Diana, G. B. Bagci, and M. Esposito, Finite-time erasing of information stored in fermionic bits, *Phys. Rev. E* **87**, 012111 (2013).
 - [6] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, Experimental verification of Landauer's principle linking information and thermodynamics, *Nature* **483**, 187 (2012).
 - [7] A. Bérut, A. Petrosyan, and S. Ciliberto, Detailed Jarzynski equality applied to a logically irreversible procedure, *Europhys. Lett.* **103**, 60002 (2013).
 - [8] Y. Jun, M. Gavrilov, and J. Bechhoefer, High-Precision Test of Landauer's Principle in a Feedback Trap, *Phys. Rev. Lett.* **113**, 190601 (2014).
 - [9] A. Bérut, A. Petrosyan, and S. Ciliberto, Information and Thermodynamics: Experimental Verification of Landauer's Erasure Principle, *J. Stat. Mech.* **2015**, P06015 (2015).
 - [10] M. Gavrilov and J. Bechhoefer, Erasure Without Work in an Asymmetric Double-Well Potential, *Phys. Rev. Lett.* **117**, 200601 (2016).
 - [11] M. Gavrilov, R. Chétrite, and J. Bechhoefer, Direct measurement of weakly nonequilibrium system entropy is consistent with Gibbs-Shannon form, *Proc. Natl. Acad. Sci. USA* **114**, 11097 (2017).
 - [12] P. R. Zulkowski and M. R. DeWeese, Optimal finite-time erasure of a classical bit, *Phys. Rev. E* **89**, 052140 (2014).
 - [13] P. R. Zulkowski and M. R. DeWeese, Optimal control of overdamped systems, *Phys. Rev. E* **92**, 032117 (2015).
 - [14] E. Aurell, K. Gawedzki, C. Mejía-Monasterio, R. Mohayae, and P. Muratore-Ginanneschi, Refined second law of thermodynamics for fast random processes, *J. Stat. Phys.* **147**, 487 (2012).
 - [15] 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](https://arxiv.org/abs/1812.11241) (2018).
 - [16] K. Proesmans, J. Ehrich, and J. Bechhoefer, Finite-time Landauer principle *Phys. Rev. Lett.* **125**, 100602 (2020).
 - [17] S. Sheng, T. Herpich, G. Diana, and M. Esposito, Thermodynamics of majority-logic decoding in information erasure, *Entropy* **21**, 284 (2019).
 - [18] E. Aurell, C. Mejía-Monasterio, and P. Muratore-Ginanneschi, Optimal Protocols and Optimal Transport in Stochastic Thermodynamics, *Phys. Rev. Lett.* **106**, 250601 (2011).

- [19] Y. Zhang, Work needed to drive a thermodynamic system between two distributions, *Europhys. Lett.* **128**, 30002 (2019).
- [20] Y. Zhang, Optimization of stochastic thermodynamic machines, *J. Stat. Phys.* **178**, 1336 (2020).
- [21] M. Esposito and C. Van den Broeck, Second law and Landauer principle far from equilibrium, *Europhys. Lett.* **95**, 40004 (2011).
- [22] C. Van den Broeck and M. Esposito, Three faces of the second law. II. Fokker-Planck formulation, *Phys. Rev. E* **82**, 011144 (2010).
- [23] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, *Rep. Prog. Phys.* **75**, 126001 (2012).
- [24] A. Dechant and Y. Sakurai, Thermodynamic interpretation of Wasserstein distance, [arXiv:1912.08405](https://arxiv.org/abs/1912.08405) (2019).
- [25] C. Browne, A. J. P. Garner, O. C. O. Dahlsten, and V. Vedral, Guaranteed Energy-Efficient Bit Reset in Finite Time, *Phys. Rev. Lett.* **113**, 100603 (2014).
- [26] M. Gopalkrishnan, A cost/speed/reliability tradeoff to erasing, *Entropy* **18**, 165 (2016).
- [27] A. P. Solon and J. M. Horowitz, Phase Transition in Protocols Minimizing Work Fluctuations, *Phys. Rev. Lett.* **120**, 180605 (2018).
- [28] O. Maillet, P. A. Erdman, V. Cavina, B. Bhandari, E. T. Mannila, J. T. Peltonen, A. Mari, F. Taddei, C. Jarzynski, V. Giovannetti, and J. P. Pekola, Optimal Probabilistic Work Extraction Beyond the Free Energy Difference with a Single-Electron Device, *Phys. Rev. Lett.* **122**, 150604 (2019).
- [29] D. H. Wolpert, C. Kempes, P. F. Stadler, and J. A. Grochow, eds., *The Energetics of Computing in Life & Machines* (Santa Fe Institute Press, Santa Fe, NM, 2019).
- [30] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, Cambridge, UK, 2007), chap. 18 .
- [31] <https://github.com/KarelProesmans/Optimal-bit-erasure/tree/master>.