Erasing a majority-logic bit

Peer-reviewed author version

epl draft

Erasing a majority-logic bit

KAREL PROESMANS[1,2,3] and JOHN BECHHOEFER[3]

[1] *Department of Physics, University of Luxembourg, Luxembourg*
[2] *Hasselt University, B-3590 Diepenbeek, Belgium*
[3] *Department of Physics, Simon Fraser University, Burnaby, B.C., V5A 1S6, Canada*

Abstract –We study finite-time bit erasure in the context of majority-logic decoding. In particular, we calculate the minimum amount of work needed to erase a majority-logic bit when one has full control over the system dynamics. We generally show that although a single unit bit is easier to erase in the slow-driving limit, the majority-logic bit outperforms the single unit bit in the fast-erasure limit. Our results also suggest optimal design principles for majority-logic bits under limited control.

Introduction. – In his seminal 1961 paper, Landauer showed that there is a fundamental thermodynamic cost associated with the erasure of a bit [1]. In particular, the amount of work needed to erase a bit is always greater than $k_B T \ln 2$, where k_B is the Boltzmann constant and T the temperature of the environment. Although modern computers still dissipate several orders of magnitude more than the Landauer limit, Koomey's law predicts that computers working near the Landauer limit will become available over the next couple of decades [2]. This raises questions about how to optimize the erasure of bits and whether there are any thermodynamic bounds on bit erasure beyond the Landauer limit.

The last two decades have seen a surge in studies on the thermodynamics of bit erasure. On the theoretical side, the framework of stochastic thermodynamics [3,4] has enabled the analysis of specific models and led to general predictions concerning the thermodynamics of finite-time bit erasure [5–15]. Meanwhile, Landauer's principle has been verified on a broad class of experimental systems [16–26]. One general prediction is that the extra work needed to erase a bit over a finite amount of time is inversely proportional to the duration of the protocol, where the proportionality constant depends on the level of control that one has over the system [6,8,9,14,15,27,28].

The focus of these works have generally been on single-unit (SU) bits, i.e., systems where the state of the bit is determined by a single spin or the position of a single colloidal particle. Realistic computers, however, are generally based on majority logic (ML) decoding [29]. An ML bit consists of several sub-bits, each of which can be either in state 0 or 1. The ML bit is then in state 0 or 1 if the majority of sub-bits is in state 0 or 1 respectively.

Inspired by a case study by Sheng et al. [30], we present here a thorough analysis of the thermodynamics of majority-logic decoding. To do so, we consider an ML bit with a large number of sub-bits under full control. We will show that SU bits are easier to erase in the slow-erasure limit but that ML bits are easier in the fast-erasure limit. In particular, the work required to erase an ML bit goes as $\tau^{-1/2}$ for fast protocols of duration τ, while an SU bit requires an amount of work proportional to τ^{-1} under similar conditions.

Stochastic thermodynamics of bit erasure . – An SU bit can generally be described by some microscopic variable x. One can, for example, think about a superconducting flux, the position of a particle or the magnetization of a spin. We define the bit to be in state 0 if $x < 0$ and in state 1 if $x > 0$. Throughout this paper, we will assume that x is a stochastic quantity that can be described by a probability distribution $p(x,t)$ at time t, evolving via a one-dimensional, overdamped Fokker-Planck equation,

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

Here, $V(x,t)$ is the potential energy landscape, D is the diffusion coefficient associated with x, k_B is the Boltzmann

constant and T is the temperature of the environment. Although we will focus in this paper on one-dimensional overdamped systems, we expect similar results to hold for underdamped and higher-dimensional systems. The only constraints on the potential is that it is initially symmetric, $V(-x,0) = V(x,0)$. Furthermore, we assume that the system is initially in equilibrium,

$$p(x,0) = \pi(x) \equiv \frac{e^{-\frac{V(x,0)}{k_B T}}}{\int_{-\infty}^{\infty} dy\, e^{-\frac{V(y,0)}{k_B T}}} \,. \quad (2)$$

Thus, at $t = 0$, the probability for the bit to be in state 0, P_0, equals the probability for it to be in state 1, P_1.

Next, we review the design of protocols that erase an SU bit to state 0 over an amount of time τ with an erasure error ϵ [15]; i.e., the probability for the bit to be in the wrong state is ϵ. The final distribution $p_\tau(x)$ should then satisfy $P_0 \equiv p(x < 0, \tau) = 1 - \epsilon$ and $P_1 \equiv p(x > 0, \tau) = \epsilon$. To find such protocols, we assume we have full control over the potential $V(x,t)$ for $0 < t < \tau$. The time-dependent function $V(x,t)$ then defines the protocol. As a further requirement, the potential-energy landscape should return to its original form at the end of the (cyclic) protocol, $V(x,\tau) = V(x,0)$. The work, on average, needed to complete this process has two contributions,

$$W = \Delta\mathcal{F} + T\Delta_i S \,, \quad (3)$$

where $\Delta\mathcal{F}$ is the non-equilibrium free energy difference between $p_0(x)$ and $p_\tau(x)$, and $\Delta_i S$ is the amount of entropy produced during the protocol.

Using ideas from stochastic thermodynamics and requiring cyclic protocols, one can show that $\Delta\mathcal{F}$ is given by the Kullback–Leibler divergence between the initial and final states [31],

$$\Delta\mathcal{F} = k_B T \int_{-\infty}^{\infty} dx\, p(x,\tau) \ln \frac{p(x,\tau)}{p(x,0)} \,. \quad (4)$$

The entropy production $\Delta_i S$ to transform a state $p(x,0)$ to a state $p(x,\tau)$ in a given time τ generally depends on the protocol, $V(x,t)$. Its minimum value, however, can be shown to be given by [6, 14, 15, 27, 32, 33]

$$\Delta_i S_{\text{min;SU}} = k_B \frac{\int_0^1 dy\, \left(f_0^{-1}(y) - f_\tau^{-1}(y)\right)^2}{D\tau} \,, \quad (5)$$

with $y = f_{0/\tau}(x) = \int_{-\infty}^{x} dx'\, p_{0/\tau}(x')$ the cumulative distribution function of the initial/final probability distribution.

Eq. (3), together with Eqs. (4)–(5) gives the least average work to erase a bit given initial and final distributions. One can further reduce the required work by relaxing the assumption that $p_\tau(x)$ be specified in advance. It is enough that $p_\tau(x)$ satisfy $f_\tau(0) = 1 - \epsilon$. Although minimizing W over these allowed final distributions cannot generally be done analytically, one can show that in
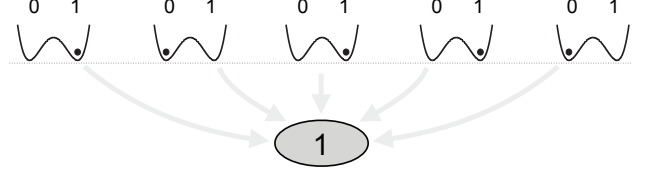

Fig. 1: Schematic diagram illustrating how $N = 5$ sub-bits combine to form one majority-logic (ML) bit.

the fast-erasure limit, $\tau \to 0$, the minimum work $W_{\text{min;SU}}$ is [14]

$$W_{\text{min;SU}} = \frac{k_B T}{D\tau} \int_0^{f_0^{-1}(1-\epsilon)} dx\, p_0(x) x^2 + o\left(\frac{1}{\tau}\right), \quad (6)$$

while in the slow-erasure limit

$$W_{\text{min;SU}} = k_B T \left[\epsilon \ln(2\epsilon) + (1 - \epsilon) \ln(2(1 - \epsilon))\right] + \mathcal{O}\left(\frac{1}{\tau}\right). \quad (7)$$

Note that Eq. (7) reduces to the Landauer bound in the zero-error limit, $\epsilon = 0$.

Majority-logic decoding. – Figure 1 shows how the state of an ML bit is determined by the sum of N sub-bits. We will assume that the sub-bits do not interact with each other and that each can be described as an SU bit that obeys Eq. (1). Since all sub-bits are equivalent, the ML bit is in state 0 if more than half of the sub-bits are in state 0 and in state 1 if more than half are in state 1. Throughout this paper, we will assume that N is odd, to avoid the possibility of a tie in deciding which state has a majority. If the probability for each sub-bit to be in state 0 is $1/2 + \delta$, one can show that the ML bit is in state 0 with probability [30]

$$P_0 = \frac{1}{2} + \frac{\int_0^{\frac{1}{2}+\delta} dx\, (x(1-x))^{\frac{N-1}{2}}}{\int_0^1 dx\, (x(1-x))^{\frac{N-1}{2}}} \,. \quad (8)$$

We show in the appendix that, for large N and letting $\sqrt{N}\delta \sim \mathcal{O}(1)$, a saddle-point approximation gives

$$P_0 \approx \frac{1}{2}\left(1 + \text{erf}(\sqrt{2N}\delta)\right). \quad (9)$$

Thus, to erase an ML bit with an erasure error ϵ, we need to erase all sub-bits by an amount δ, given by

$$\delta \approx \frac{\text{erf}^{-1}(1 - 2\epsilon)}{\sqrt{2N}} \,. \quad (10)$$

Note that $\delta = 0$ corresponds to no erasure, as the probabilities to be in state 0 and 1 remain equal.

The total work associated with the erasure of an ML bit with a large number of sub-bits now equals N times the thermodynamic erasure cost of a single bit with erasure error $1/2 - \delta$. To calculate the work associated with the

erasure of an ML bit, we first introduce the coordinate transformation

$$\Gamma_{\text{sb}}(x) = f_{\text{sb},\tau}^{-1}(f_{\text{sb},0}(x)), \tag{11}$$

where $f_{\text{sb},0/\tau}(x)$ is the cumulative distribution associated with the state of the sub-bit at time 0 or τ, $p_{\text{sb},0/\tau}(x)$, respectively. Within the mathematical framework of optimal transport theory, $\Gamma(x)$ is known as the *transport map*. Physically, it can be interpreted as the position to which one moves the probability density that was originally at position x. If $\delta = 0$, the final macroscopic state of each sub-bit bit equals its initial state, $P_0 = P_1 = 1/2$, and the final microscopic state of the sub-bits is the same as the distribution $p_{\text{sb},\tau}(x) = p_{\text{sb},0}(x)$. This implies that $\Gamma_{\text{sb}}(x) = x$. Therefore, in the limit of many sub-bits, one can Taylor expand $\Gamma_{\text{sb}}(x)$ to first order,

$$\Gamma_{\text{sb}}(x) = x + \frac{\Gamma_1(x)}{\sqrt{N}} + \mathcal{O}\left(\frac{1}{N}\right), \tag{12}$$

where $\Gamma_1(x)$ depends on the erasure protocol and can be interpreted physically as the total distance that the probability density originally at position x is transported by time τ. In the appendix, we use Eq. (3) to show that one can now write

$$W_{\text{min};\text{ML}} = N W_{\text{min};\text{sb}}$$
$$\approx \int_{-\infty}^{\infty} dx \left(\frac{(p_0(x)\Gamma_1'(x) + p_0'(x)\Gamma_1(x))^2}{2p_0(x)} \right.$$
$$\left. + \frac{p_0(x)\Gamma_1(x)^2}{D\tau} \right). \tag{13}$$

Furthermore, the constraint $f_{\text{sb},\tau}(0) = 1/2 + \delta$ can be rewritten as

$$\Gamma_1(0) = -\frac{\text{erf}^{-1}(1 - 2\epsilon)}{\sqrt{2}p_0(0)}. \tag{14}$$

To calculate the minimum average work needed to erase an ML bit with erasure error ϵ, one can minimize Eq. (13) over all $\Gamma_1(x)$ that satisfy Eq. (14). Lagrange's method of constrained optimization then leads to,

$$\frac{d}{dx} \frac{(V'(x)\Gamma_1(x))}{k_{\text{B}}T} - \frac{d^2}{dx^2}\Gamma_1(x) + \frac{2}{D\tau}\Gamma_1(x) = \lambda\delta(x), \tag{15}$$

where λ is a Lagrange multiplier. As Eq. (15) is a linear differential equation, once the initial potential-energy landscape of the sub-bits, $V(x,0)$, is specified, one can in principle determine analytically $\Gamma_1(x)$ and therefore the minimum work to erase an ML bit with many sub-bits.

In the slow- and fast-erasure limits, one can calculate $W_{\text{min};\text{ML}}$ exactly, even without specifying the $V(x,0)$. Indeed, in the slow-erasure limit, the work needed to erase the bit just reduces to the total change in Shannon entropy,

$$W_{\text{min};\text{ML}} \approx k_{\text{B}}T(\text{erf}^{-1}(1 - 2\epsilon))^2. \tag{16}$$
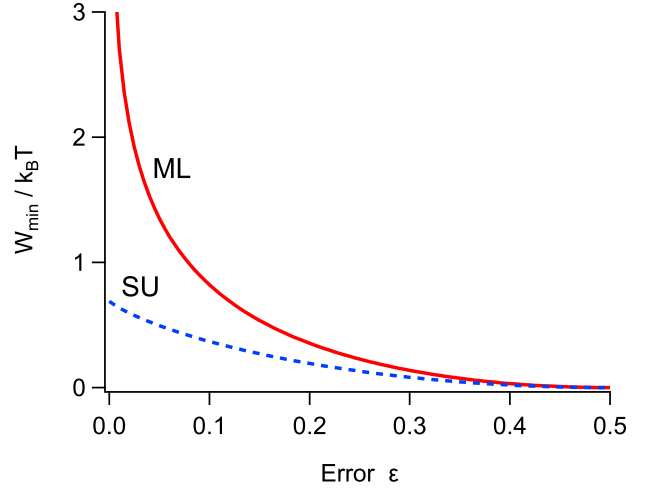

Fig. 2: Minimum average work W_{min} for a single-unit bit (blue dashed line) and a majority-logic bit (solid red line) in the slow-erasure limit.

In Fig. 2, we compare the slow-erasure costs of SU bit, Eq. (7) with that of an ML bit, Eq. (16). One can immediately see that it always takes more work to erase an ML bit than an SU bit, whatever the erasure error. In the fast-erasure limit, a WKB calculation leads to

$$W_{\text{min};\text{ML}} \approx \frac{k_{\text{B}}T(\text{erf}^{-1}(1 - 2\epsilon))^2}{p_0(0)\sqrt{2D\tau}}. \tag{17}$$

In the fast-erasure limit, the work needed to erase an ML bit scales as $\tau^{-1/2}$; by contrast, this cost for an SU bit scales as τ^{-1}, c.f., Eq. (6). Therefore, in the fast-erasure limit, the cost associated with an SU bit scales quadratically worse than the cost associated with an ML bit.

Finite-N effects. To arrive at the above results, Eq. 15, we first took the limit $N \to \infty$ and then the limit $\tau \to 0$. Any realistic ML bit does, however, have a finite number of sub-bits. Therefore, for very fast erasure, the limit $\tau \to 0$ becomes the dominant limit and should be taken prior to the $N \to \infty$ limit. Using Eq. (6), one then arrives at

$$W_{\text{min};\text{ML}} \approx \frac{k_{\text{B}}T(\text{erf}^{-1}(1 - 2\epsilon))^3}{\sqrt{72N}p_0(0)^2 D\tau}. \tag{18}$$

Comparing this result with Eq. (17), one can see that the finite-N scaling becomes relevant for $\tau \sim \left(Dp_0(0)^2 N\right)^{-1}$. Furthermore, by comparing Eq. (18) with Eq. (6), one can see that the ML bit can still outperform an SU bit by a factor $\sqrt{N}$.

Example: Flat well . – Let us now apply our framework to a simple example, where we assume a large number of sub-bits ($N \to \infty$) and where the initial potential has a flat bottom of width $2L$ and infinite walls,

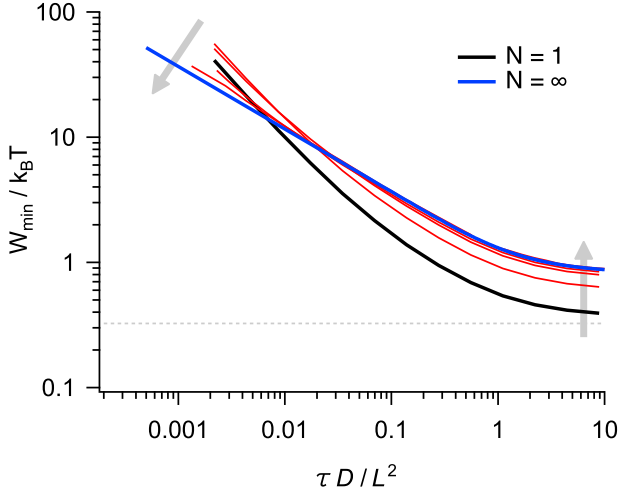$$V(x,0) = \begin{cases} 0 & -L < x < L \\ \infty & \text{otherwise} \end{cases}. \tag{19}$$

Fig. 3: Minimum average work $W_{\mathrm{min;ML}}$ to erase an ML bit as a function of scaled protocol duration $\tau' = \tau D/L^2$, for ML bits with $\epsilon = 0.1$. The thick black curve corresponds to $N = 1$ sub-bit. The thin red curves show $N = 3, 9, 33, 99$, and the thick black curve is the $N \to \infty$ result given in Eq. (3). Light-shaded arrows show trends as N is increased from 1 to ∞.

To calculate the minimum work needed to erase the bit, we first need to find $\Gamma_1(x)$ from Eq. (15), with $V(x,0)$ given by Eq. (19). From Eqs. (11) and (12), it is clear that the solution should satisfy the boundary conditions

$$\Gamma_1(-L) = \Gamma_1(L) = 0. \qquad (20)$$

Together with the constraint provided by Eq. (14), this leads to the solution (see appendix)

$$\Gamma_1(x) = \begin{cases} -\dfrac{\sqrt{2}L\,\mathrm{erf}(1-2\epsilon)\sinh\left(\sqrt{\frac{2}{\tau'}}\left(\frac{x}{L}+1\right)\right)}{\sinh\left(\sqrt{\frac{2}{D\tau}}L\right)} & x < 0, \\[3mm] -\dfrac{\sqrt{2}L\,\mathrm{erf}(1-2\epsilon)\sinh\left(\sqrt{\frac{2}{\tau'}}\left(\left(\frac{x}{L}-1\right)\right)\right)}{\sinh\left(-\sqrt{\frac{2}{\tau'}}\right)} & x > 0. \end{cases} \qquad (21)$$

Plugging this expression for $\Gamma_1(x)$ into Eq. (13) then gives us the minimum work to erase an ML bit made up of a large number of sub-bits that each have an initial flat-well potential described by Eq. (19):

$$W_{\mathrm{min;ML}} = \frac{k_{\mathrm{B}}T\,(\mathrm{erf}^{-1}(1-2\epsilon))^2}{\sqrt{2\tau'}}\frac{\sinh\left(2\sqrt{\frac{2}{\tau'}}\right)}{\cosh^2\left(\sqrt{\frac{2}{\tau'}}\right)-1}, \qquad (22)$$

with $\tau' \equiv \tau D/L^2$ the scaled protocol time.

In Fig. 3, we plot the minimum work needed to erase an ML bit, for various numbers of sub-bits. For large τ, the cost to erase an SU bit is lower than the cost to erase an ML bit; however, for $\tau \lesssim \mathrm{Var}(x)/D$, ML bits are easier to erase. In other words, an optimal protocol would use a single unit for slow erasure and majority logic for fast erasure. This conclusion that the optimal strategy shifts suddenly as a function of a parameter (the number

of sub-bits N) recalls the scenarios of phase-transitions in optimal protocols studied in [34]. One can also see that for very small τ, the finite-size effects of ML bits become important, making them more costly to erase.

Discussion. – We have shown, for a broad class of potentials (i.e., all symmetric potentials with a stable equilibrium state), that erasing a majority-logic bit is more efficient for sufficiently fast erasure protocols. What is the origin of this behaviour? In an ML bit, one neglects some microscopic degrees of freedom. As a result, both the entropy and free-energy difference between the initial and final states of an ML bit exceed those of an SU bit. For fast erasure, the dissipative part of the work needed to erase the bit dominates. Following Eq. (21) and interpreting $\Gamma_{\mathrm{sb},1}(x)$ as the displacement of the probability distribution, one can see that optimal fast erasure for ML bits consists of changing the states only for sub-bits where $x \approx 0$. Meanwhile, sub-bits that have $x \gg 0$ and are therefore harder to erase remain untouched. The lower cost of sub-bits with $x \approx 0$ also explains why the erasure cost of an ML bit is inversely proportional to the probability that $x \approx 0$, as suggested by Eq. (17).

This conclusion points to an important design principle for realistic ML bits: Generally, one does not have full control over the energy landscape of the sub-bits, and control is limited to the possibilities provided by a finite number of parameters. Limiting the space of control algorithms will increase the minimum work given by Eq. (17). The above argument does, however, show that the main ingredient to reduce the operational cost of an ML bit is the possibility to modify the energy landscape around $x = 0$ without modifying it away from this region. That kind of manipulation of the potential can be created using a limited set of control parameters.

It is worth noting that other effects might also lead to sublinear scaling of the work cost in the fast-erasure limit. For example, Pancotti et al. have recently suggested that a time-dependent system-bath coupling might lead to similar effects [35]. Furthermore, effects inside the bath, such as viscoelasticity, might influence the erasure cost [32]. These mechanisms are quite distinct from the one proposed here, which has the advantage of applying to the types of memories used in technological applications, where a single bit consists of many spins and the average magnetization records the bit of information.

Can one extend our analysis to other types of systems? Optimal transport theory predicts that the minimum dissipation to transform a microscopic system from an initial to a final state is inversely proportional to the duration of the protocol. Here, we showed that if the mapping from microscopic to macroscopic dynamics is non-trivial, this conclusion need not hold for macroscopic transformations. Similar analyses might then be possible in other settings, such as chemical reaction networks or heat engines.

* * *

Appendix: erasure error. – Here, we find the erasure error of an ML bit given the erasure errors of its sub-bits. First, we calculate the probability, P_0, that an MLB with N sub-bit is in state 0 when all of the sub-bits are in state 0 with probability p.

It was shown in [30] that

$$P_0 = \frac{\int_0^p \mathrm{d}t\,(t(1-t))^{\frac{N-1}{2}}}{\int_0^1 \mathrm{d}t\,(t(1-t))^{\frac{N-1}{2}}}. \tag{23}$$

Let us now look at the large-N limit, where the integrands in both the numerator and denominator of Eq. (23) are exponentially small unless $t \approx 1/2$. We can then do a saddle-point approximation:

$$P_0 \approx \frac{\int_0^p \mathrm{d}t\,\mathrm{e}^{-N\left(\ln 2 + 4\left(t-\frac{1}{2}\right)^2\right)}}{\int_0^1 \mathrm{d}t\,\mathrm{e}^{-N\left(\ln 2 + 4\left(t-\frac{1}{2}\right)^2\right)}}$$

$$\approx \frac{1}{2}\left[1 + \mathrm{erf}\left(\sqrt{N}(2p-1)\right)\right], \tag{24}$$

where we used the Taylor expansion of the log term around $t = 1/2$,

$$\ln(t(1-t)) \approx -2\ln 2 - 4\left(t - \frac{1}{2}\right)^2 + \mathcal{O}\left(\left(t - \frac{1}{2}\right)^4\right) \tag{25}$$

to get the second equality. Note that when $\sqrt{N}(p-1/2) \ll 0$ or $\sqrt{N}(p-1/2) \gg 0$, one gets $P_0 \approx 0$ or $P_0 \approx 1$ respectively, up to an exponentially small correction.

Appendix: optimal work. – Let us look at the large-N limit and assume that $\Gamma_{\mathrm{sb}}(x)$ is of the form

$$\Gamma_{\mathrm{sb}}(x) = x + \frac{\Gamma_1(x)}{\sqrt{N}} + \mathcal{O}\left(\frac{1}{N}\right). \tag{26}$$

The dissipation associated with the erasure of one sub-bit is given by

$$\Delta_{\mathrm{i}}S_{\mathrm{min;sb}} = \frac{\int_0^1 \mathrm{d}y\,\left(f_{\mathrm{sb},0}^{-1}(y) - f_{\mathrm{sb},\tau}^{-1}(y)\right)}{D\tau}$$

$$= \frac{\int_{-\infty}^\infty \mathrm{d}x\,p_0(x)\,(\Gamma_{\mathrm{sb}}(x) - x)}{D\tau}$$

$$\approx \int_{-\infty}^\infty \mathrm{d}x\,\frac{p_0(x)\Gamma_1(x)^2}{ND\tau}, \tag{27}$$

where we changed variables from $x \to f_0(x)$ in the second line. For the nonequilibrium free-energy difference, we first note that as $f_{\mathrm{sb},0}(x) = f_{\mathrm{sb},\tau}(\Gamma_{\mathrm{sb}}(x))$, we can write,

$$p_{\mathrm{sb},0}(x) = \Gamma_{\mathrm{sb}}'(x)p_{\mathrm{sb},\tau}(\Gamma_{\mathrm{sb}}(x)). \tag{28}$$

This leads to

$$\frac{\Delta\mathcal{F}}{k_{\mathrm{B}}T} = \int_{-\infty}^\infty \mathrm{d}x\,p_0(x)\ln\frac{p_0(x)}{\Gamma'(x)p_0(\Gamma(x))}, \tag{29}$$

where we changed variables, $x \to \Gamma(x)$ in the second line. Plugging in Eq. (26) and doing a lowest-order expansion in $1/\sqrt{N}$ leads to

$$\frac{\Delta\mathcal{F}}{k_{\mathrm{B}}T} = \frac{1}{N}\int_{-\infty}^\infty \mathrm{d}x\left(\frac{\left(\frac{\mathrm{d}}{\mathrm{d}x}(p_0(x)\Gamma_1(x))\right)^2}{2p_0(x)}\right), \tag{30}$$

and therefore we arrive at

$$\frac{W_{\mathrm{min;ML}}}{k_{\mathrm{B}}T} = \int_{-\infty}^\infty \mathrm{d}x\left(\frac{\left(\frac{\mathrm{d}}{\mathrm{d}x}(p_0(x)\Gamma_1(x))\right)^2}{2p_0(x)} + \frac{p_0(x)\Gamma_1(x)^2}{D\tau}\right). \tag{31}$$

Now let us reformulate the constraint

$$f_{\mathrm{sb},\tau}(0) = \frac{1}{2} + \delta, \quad \text{or} \quad \Gamma_{\mathrm{sb}}\left(f_{\mathrm{sb},0}^{-1}\left(\frac{1}{2} + \delta\right)\right) = 0 \tag{32}$$

in terms of $\Gamma_1(x)$. First, we note that $f_0^{-1}(1/2) = 0$. We then use a first-order Taylor expansion and Eq. (24) to show that

$$f_{\mathrm{sb},0}^{-1}\left(\frac{1}{2} + \delta\right) = \frac{\delta}{f_{\mathrm{sb},0}'\left(f_{\mathrm{sb},0}^{-1}\left(\frac{1}{2}\right)\right)} + O(\delta^2)$$

$$= \frac{\mathrm{erf}^{-1}(1-2\epsilon)}{p_0(0)\sqrt{2N}} + \mathcal{O}\left(\frac{1}{N}\right). \tag{33}$$

This leads to

$$\Gamma_{\mathrm{sb}}\left(f_{\mathrm{sb},0}^{-1}\left(\frac{1}{2} + \delta\right)\right) \approx \frac{\mathrm{erf}^{-1}(1-2\epsilon)}{p_0(0)\sqrt{2N}} + \frac{\Gamma_1(0)}{\sqrt{N}} = 0, \tag{34}$$

or

$$\Gamma_1(0) = -\frac{\mathrm{erf}^{-1}(1-2\epsilon)}{\sqrt{2}p_0(0)}, \tag{35}$$

which can be rewritten to

$$\int_{-\infty}^\infty \mathrm{d}x\,\Gamma_1(x)\delta(x) = -\frac{\mathrm{erf}^{-1}(1-2\epsilon)}{\sqrt{2}p_0(0)}. \tag{36}$$

The minimization of Eq. (31) under the constraint of Eq. (36) can now be done by introducing the Lagrangian

$$\mathcal{L} = \int_{-\infty}^\infty \mathrm{d}x\left(\frac{\left(\frac{\mathrm{d}}{\mathrm{d}x}(p_0(x)\Gamma_1(x))\right)^2}{2p_0(x)} + \frac{p_0(x)\Gamma_1(x)^2}{D\tau}\right)$$

$$+ \lambda\left(\int_{-\infty}^\infty \mathrm{d}x\,\Gamma_1(x)\delta(x) + \frac{\mathrm{erf}^{-1}(1-2\epsilon)}{\sqrt{2}p_0(0)}\right) \tag{37}$$

and solving the Euler-Lagrange equation

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{\delta\mathcal{L}}{\delta\Gamma_1'(x)} = \frac{\delta\mathcal{L}}{\delta\Gamma_1(x)}, \tag{38}$$

which leads to

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(p_0(x)\Gamma_1'(x)\right) + p_0''(x)\Gamma_1(x) - \frac{p_0'(x)^2\Gamma_1(x)}{p_0(x)}$$
$$- \frac{2p_0(x)\Gamma_1(x)}{D\tau} = \lambda\delta(x), \quad (39)$$

or using the fact that the system was initially in a Boltzmann distribution,

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{(\Gamma_1(x)V'(x))}{k_\mathrm{B}T} - \frac{\mathrm{d}^2}{\mathrm{d}x^2}\Gamma_1(x) + \frac{2\Gamma_1(x)}{D\tau} = \lambda\delta(x) \quad (40)$$

Appendix: slow- and fast-erasure limits. – For slow erasure ($\tau \to \infty$), the minimum work needed to erase a bit is dominated by the free-energy difference between the initial and final states of the system,

$$\frac{W_{\mathrm{min;ML}}}{k_\mathrm{B}T} \approx \frac{\Delta F}{k_\mathrm{B}T}$$
$$= -N\left[\epsilon_{\mathrm{sb}}\ln\epsilon_{\mathrm{sb}} + (1-\epsilon_{\mathrm{sb}})\ln(1-\epsilon_{\mathrm{sb}}) - \ln 2\right], \quad (41)$$

where the second line follows from the fact that the local equilibrium distribution,

$$p_\tau(x) = \begin{cases} 2(1-\epsilon)p_0(x) & x < 0 \\ 2\epsilon p_0(x) & x > 0 \end{cases} \quad (42)$$

minimizes the free-energy difference between the initial and final states [15]. Using, $\epsilon_{\mathrm{sb}} = 1/2 - \delta$ with δ small, we can rewrite this as

$$W_{\mathrm{min;ML}} \approx 2Nk_\mathrm{B}T\delta^2$$
$$= k_\mathrm{B}T\,\mathrm{erf}^{-1}\left(1-2\epsilon\right)^2. \quad (43)$$

For fast erasure ($\tau \to 0$), we first need to solve Eq. (40). To do this, we first make a WKB ansatz for $\Gamma_1(x)$,

$$\Gamma_1(x) \sim \exp\left(\frac{\alpha_0(x)}{\sqrt{\tau}} + \alpha_1(x) + O(\sqrt{\tau})\right). \quad (44)$$

Eq. (40) then becomes

$$\frac{2 - D\alpha_0'(x)^2}{D\tau} + \frac{a_0'(x)\left(\frac{V'(x)}{k_\mathrm{B}T} - 2a_1'(x)\right) - a_0''(x)}{\sqrt{\tau}}$$
$$+ O(1) = \lambda\delta(x). \quad (45)$$

We can now approximate $\Gamma_1(x)$ by solving the above equation order-by-order in τ:

$$2 - D\alpha_0'(x)^2 = 0 \quad (46)$$
$$a_0'(x)\left(V'(x) - 2a_1'(x)\right) - a_0''(x) = 0. \quad (47)$$

This leads to

$$\alpha_0(x) = \pm\sqrt{\frac{2}{D}}x, \qquad \alpha_1(x) = \frac{V(x)}{2k_\mathrm{B}T}. \quad (48)$$

Finally, as Eq. (14) is a homogeneous linear differential equation for $x \neq 0$, any linear combination of solutions is itself a solution. The delta function has the effect that the first derivative of $\Gamma_1(x)$ need not be continuous at $x \neq 0$. Therefore, the general solution can be written as

$$\Gamma_1(x) = \begin{cases} \mathrm{e}^{\frac{V(x)}{2k_\mathrm{B}T}}\left(C_{0,-}\,\mathrm{e}^{\sqrt{\frac{2}{D\tau}}x} + C_{1,-}\,\mathrm{e}^{-\sqrt{\frac{2}{D\tau}}x}\right) & x < 0, \\ \mathrm{e}^{\frac{V(x)}{2k_\mathrm{B}T}}\left(C_{0,+}\,\mathrm{e}^{\sqrt{\frac{2}{D\tau}}x} + C_{1,+}\,\mathrm{e}^{-\sqrt{\frac{2}{D\tau}}x}\right) & x > 0. \end{cases}$$
$$(49)$$

One can now fix $C_{0,-}$, $C_{1,-}$, $C_{0,+}$ and $C_{1,+}$ by using the boundary conditions, Eq. (35), and $\Gamma_1(x_{\mathrm{min}}) = \Gamma_1(x_{\mathrm{max}}) = 0$. Using these conditions, we can write

$$\Gamma_1(x) = \begin{cases} \frac{\mathrm{erf}^{-1}(1-2\epsilon)}{\sqrt{2}p_0(0)}\mathrm{e}^{\frac{V(x)-V(0)}{2k_\mathrm{B}T}}\frac{\sinh\left(\sqrt{\frac{2}{D\tau}}(x-x_{\mathrm{min}})\right)}{\sinh\left(\sqrt{\frac{2}{D\tau}}x_{\mathrm{min}}\right)} & x < 0, \\ \frac{\mathrm{erf}^{-1}(1-2\epsilon)}{\sqrt{2}p_0(0)}\mathrm{e}^{\frac{V(x)-V(0)}{2k_\mathrm{B}T}}\frac{\sinh\left(\sqrt{\frac{2}{D\tau}}(x-x_{\mathrm{max}})\right)}{\sinh\left(\sqrt{\frac{2}{D\tau}}x_{\mathrm{max}}\right)} & x > 0. \end{cases}$$
$$(50)$$

For small τ, it is clear that $\Gamma_1(x)$ is exponentially small unless $|x| \approx 0$, where one can write

$$\Gamma_1(x) \approx \frac{\mathrm{erf}^{-1}(1-2\epsilon)}{\sqrt{2}p_0(0)}\mathrm{e}^{-\sqrt{\frac{2}{D\tau}}|x|}. \quad (51)$$

Therefore, we can do a saddle-point-like approximation on Eq. (31), which leads to

$$\frac{W_{\mathrm{min;ML}}}{k_\mathrm{B}T} \approx \left(\int_{-\infty}^{\infty}\mathrm{d}x\,\frac{\left(p_0'(0)\,\mathrm{e}^{-\sqrt{\frac{2}{D\tau}}|x|} - p_0(0)\sqrt{\frac{2}{D\tau}}\,\mathrm{e}^{-\sqrt{\frac{2}{D\tau}}|x|}\right)^2}{2p_0(0)}\right.$$
$$\left. + \int_{-\infty}^{\infty}\mathrm{d}x\,\frac{p_0(0)\,\mathrm{e}^{-2\sqrt{\frac{2}{D\tau}}|x|}}{D\tau}\right)\left(\frac{\mathrm{erf}^{-1}(1-2\epsilon)}{\sqrt{2}p_0(0)}\right)^2$$
$$\approx \frac{2\,\mathrm{erf}^{-1}(1-2\epsilon)^2}{p_0(0)D\tau}\int_{-\infty}^{\infty}\mathrm{d}x\,\mathrm{e}^{-2\sqrt{\frac{2}{D\tau}}|x|}$$
$$= \frac{\sqrt{2}\,\mathrm{erf}^{-1}(1-2\epsilon)^2}{p_0(0)\sqrt{D\tau}}. \quad (52)$$

Appendix: flat well. – As an example, let us look at a bit consisting of a particle in a flat potential well of width $2L$,

$$V(x) = \begin{cases} 0 & -L < x < L \\ \infty & \text{otherwise} \end{cases} \quad (53)$$

The equilibrium distribution is given by

$$p_0(x) = \frac{1}{2}L, \quad (54)$$

for $-L < x < L$ and zero otherwise. Eq. (14) then becomes

$$-\frac{\mathrm{d}^2}{\mathrm{d}x^2}\Gamma_1(x) + \frac{2\Gamma_1(x)}{D\tau} = \lambda\delta(x), \quad (55)$$
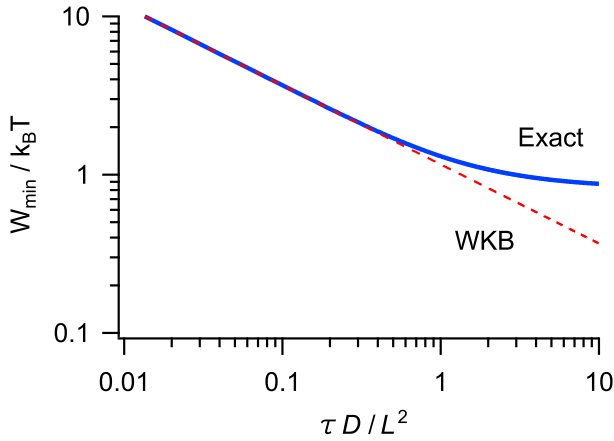
Fig. 4: Minimum average work $W_{\mathrm{min;ML}}$ to erase an ML bit as a function of scaled protocol duration $\tau' = \tau D/L^2$, for an ML bit with $\epsilon = 0.1$. The thick blue curve corresponds to the exact $N = \infty$ solution (Eq. 57). The dashed red curve shows the WKB approximation from Eq. (52), with $p_0(0) = \frac{1}{2}L$.

with boundary conditions $\Gamma_1(-L) = \Gamma_1(L) = 0$ and $\Gamma_1(0) = -\sqrt{2}l\,\mathrm{erf}^{-1}(1 - 2\epsilon)$. Therefore, one gets

$$\Gamma_1(x) = \begin{cases} \sqrt{2}L\,\mathrm{erf}^{-1}(1-2\epsilon)\dfrac{\sinh\left(\sqrt{\frac{2}{D\tau}}(x+L)\right)}{\sinh\left(\sqrt{\frac{2}{D\tau}}L\right)} & x < 0 \\[3mm] -\sqrt{2}L\,\mathrm{erf}^{-1}(1-2\epsilon)\dfrac{\sinh\left(\sqrt{\frac{2}{D\tau}}(x-L)\right)}{\sinh\left(\sqrt{\frac{2}{D\tau}}L\right)} & x > 0 \end{cases}$$
(56)

This leads to

$$\begin{aligned} W_{\mathrm{min;ML}} &= \int_{-L}^{L} dx \left(\frac{\Gamma_1'(x)^2}{4L} + \frac{\Gamma_1(x)^2}{2LD\tau} \right) \\ &= \frac{k_{\mathrm{B}}TL\left(\mathrm{erf}^{-1}(1-2\epsilon)\right)^2}{\sqrt{2D\tau}} \frac{\sinh\left(2L\sqrt{\frac{2}{D\tau}}\right)}{\left(\cosh\left(\sqrt{\frac{2}{D\tau}}L\right)\right)^2 - 1} . \end{aligned}$$
(57)

REFERENCES

[1] Landauer R., *IBM J. Res. Dev.*, **5** (1961) 183.
[2] Koomey J., Berard S., Sanchez M. and Wong H., *IEEE Ann. Hist. Comput.*, **33** (2010) 46.
[3] Seifert U., *Rep. Prog. Phys.*, **75** (2012) 126001.
[4] Van den Broeck C. and Esposito M., *Physica A*, **418** (2015) 6.
[5] Dillenschneider R. and Lutz E., *Phys. Rev. Lett.*, **102** (2009) 210601.
[6] Aurell E., Gawedzki K., Mejia-Monasterio C., Mohayaee R. and Muratore-Ginanneschi P., *J. Stat. Phys.*, **147** (2012) 487.
[7] Diana G., Bagci G. B. and Esposito M., *Phys. Rev. E*, **87** (2013) 012111.
[8] Zulkowski P. R. and DeWeese M. R., *Phys. Rev. E*, **89** (2014) 052140.
[9] Zulkowski P. R. and DeWeese M. R., *Phys. Rev. E*, **92** (2015) 032117.
[10] Deshpande A., Gopalkrishnan M., Ouldridge T. E. and Jones N. S., *Proc. R. Soc. Lond.*, **473** (2017) 20170117.
[11] Melbourne J., Talukdar S. and Salapaka M. V., *Proc. IEEE Conf. on Decision and Control (CDC)*, (2018) 4135.
[12] Boyd A. B., Patra A., Jarzynski C. and Crutchfield J. P., *arXiv:1812.11241*, (2018) .
[13] Riechers P. M., Boyd A. B., Wimsatt G. W. and Crutchfield J. P., *Phys. Rev. Research*, **2** (2020) 033524.
[14] Proesmans K., Ehrich J. and Bechhoefer J., *Phys. Rev. Lett.*, **125** (2020) 100602.
[15] Proesmans K., Ehrich J. and Bechhoefer J., *Phys. Rev. E*, **102** (2020) 032105.
[16] Bérut A., Arakelyan A., Petrosyan A., Ciliberto S., Dillenschneider R. and Lutz E., *Nature*, **483** (2012) 187.
[17] Bérut A., Petrosyan A. and Ciliberto S., *EPL*, **103** (2013) 60002.
[18] Jun Y., Gavrilov M. and Bechhoefer J., *Phys. Rev. Lett.*, **113** (2014) 190601.
[19] Gavrilov M. and Bechhoefer J., *Phys. Rev. Lett.*, **117** (2016) 200601.
[20] Hong J., Lambson B., Dhuey S. and Bokor J., *Sci. Adv.*, **2** (2016) e1501492.
[21] Martini L., Pancaldi M., Madami M., Vavassori P., Gubbiotti G., Tacchi S., Hartmann F., Emmerling M., Höfling S., Worschech L. et al., *Nano Energy*, **19** (2016) 108.
[22] Gavrilov M., Chétrite R. and Bechhoefer J., *Proc. Natl. Acad. Sci. U.S.A.*, **114** (2017) 11097.
[23] Gaudenzi R., Burzurí E., Maegawa S., van der Zant H. and Luis F., *Nat. Phys.*, **14** (2018) 565.
[24] Ribezzi-Crivellari M. and Ritort F., *Nat. Phys.*, **15** (2019) 660.
[25] Saira O.-P., Matheny M. H., Katti R., Fon W., Wimsatt G., Crutchfield J. P., Han S. and Roukes M. L., *Phys. Rev. Research*, **2** (2020) 013249.
[26] Cetiner U., Raz O. and Sukharev S., *bioRxiv*, (2020) 2020.06.26.174649;.
[27] Aurell E., Mejía-Monasterio C. and Muratore-Ginanneschi P., *Phys. Rev. Lett.*, **106** (2011) 250601.
[28] Schmiedl T. and Seifert U., *EPL*, **83** (2008) 30005.
[29] Richter H. J., *J. Phys. D*, **32** (1999) R147.
[30] Sheng S., Herpich T., Diana G. and Esposito M., *Entropy*, **21** (2019) 284.
[31] Esposito M. and Van den Broeck C., *EPL*, **95** (2011) 40004.
[32] Zhang Y., *EPL*, **128** (2019) 30002.
[33] Zhang Y., *J. Stat. Phys.*, **178** (2020) 1336.
[34] Solon A. P. and Horowitz J. M., *Phys. Rev. Lett.*, **120** (2018) 180605.
[35] Pancotti N., Scandi M., Mitchison M. T. and Perarnau-Llobet M., *Phys. Rev. X*, **10** (2020) 031015.