Designing the Optimal Bit: Balancing Energetic Cost, Speed and Reliability

Abhishek Deshpande[*1,2], Manoj Gopalkrishnan[3], Thomas E. Ouldridge[4], and Nick S. Jones[1]

[1]*Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom*
[2]*School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai 400005, India*
[3]*Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India*
[4]*Department of Bioengineering, Imperial College London, London SW7 2AZ, United Kingdom*

Abstract

We consider the challenge of operating a reliable bit that can be rapidly erased. We find that both erasing and reliability times are non-monotonic in the underlying friction, leading to a trade-off between erasing speed and bit reliability. Fast erasure is possible at the expense of low reliability at moderate friction, and high reliability comes at the expense of slow erasure in the underdamped and overdamped limits. Within a given class of bit parameters and control strategies, we define "optimal" designs of bits that meet the desired reliability and erasing time requirements with the lowest operational work cost. We find that optimal designs always saturate the bound on the erasing time requirement, but can exceed the required reliability time if critically damped. The non-trivial geometry of the reliability and erasing time-scales allows us to exclude large regions of parameter space as suboptimal. We find that optimal designs are either critically damped or close to critical damping under the erasing procedure.

Keywords: erasing/switching a bit, particle in a double well, reliability of information, optimal bit, friction trade-off, saturation/unsaturation of time-scales

1 Introduction

Certain information processing operations such as erasing a bit, or copying the state of one bit into another previously randomised bit have fundamental lower bounds on work input [1, 2, 3, 4, 5]. These lower bounds such as the famous $k_B T \ln 2$ minimal cost for erasing arise due to equilibrium thermodynamics: there is a need to compensate for any entropy reduction in the information-carrying system with an entropy increase elsewhere. Practical devices, however, do not approach these bounds [6, 7] and insights gained from thinking about the lower bound have not yet translated into more energy-efficient technology. A partial explanation is that man-made devices and biological cells need to operate on fast time-scales and hence cannot involve the quasistatic manipulations necessary to reach lower bounds [8, 9]. An alternative

---

[*]Corresponding author: deshabhi123@gmail.com

suggestion from von Neumann is that the need to store information for long periods of time (**reliability**) leads to high-cost architectures [10]. We explore the interplay between reliability, speed and the energetic cost of bit operation. Equilibrium thermodynamic bounds such as the Landauer limit cannot account for these inherently kinetic phenomena.

This general question of how to design fast, cheap and reliable bits has obvious technological relevance to the optimal design of low power computational devices [11, 12, 13] Additionally, since the discovery of the structure of DNA and the central dogma of molecular biology, it has become well accepted that information processing is at the heart of many natural phenomena. Many authors have explored information processing in biological systems, both to understand natural examples and design synthetic analogs [3, 14, 15, 16, 9, 17, 18, 19]. The question of the interplay between reliability, speed and cost are also relevant here, although under-explored.

In this paper, we explore the challenge of building fast, cheap and reliable bits, and provide a framework for it's analysis in terms of reliability and erasure time-scales. We also take the first steps towards exploring the physics of the optimal design problem by considering a simple model: a particle in a 1-D potential, which is a quartic double-well potential in the device's "resting" state. We require that the bit be reliable, so that a particle equilibrated in either well stays in that well for a specified long time on average. Simultaneously, we require the implementation of an "erase" or "reset" operation using an external control, so that erasure is completed within a specified short amount of time. Our principal question is to find values for the design parameters which consist of the height of the double well, the friction coefficient, and the control parameters to guarantee these requirements without expending more energy than required. Our main contribution is an exploration of this design space, which demonstrates the previously under-appreciated role of friction. In particular, we identify a "Goldilocks zone" where the friction coefficient takes moderate values. This is somewhat counter-intuitive because historically friction has been viewed as a nuisance to computing, to be sent as low as possible [20, 21, 22, 23].

In Section 2, we describe the model which will provide intuition for our work. We formalize the time-scale over which the bit stores information through the notion of **reliability time**. In Section 2.2.1, we describe one simple family of control protocols for resetting a bit. We calculate the work done in erasing a bit for this form of control. We will use this particular control protocol to illustrate our subsequent ideas. In Section 2.2.2, we introduce the notion of **erasing time**. In Section 3, we consolidate from the literature the analytical forms and approximations for our two time-scales of interest, and confirm them with numerical simulations. We find that both the reliability and erasing time-scales are non-monotonic, roughly U-shaped functions of the friction coefficient. It follows that high reliability is obtained by setting the friction to a low or high value, whereas a low erasing time is favoured by an intermediate value of friction, implying a conflict between the two time scales for a given class of protocols. In Section 4, we investigate how this conflict feeds into the geometry of **optimal bits**: bits that fulfil the desired reliability and erasing time requirements with the minimum energy cost. We find and partially characterize a "Goldilocks zone" in design space where optimal bits reside. In Section 5, we discuss the robustness of our results when more freedom is allowed in the choice of design parameters, and the control protocol.
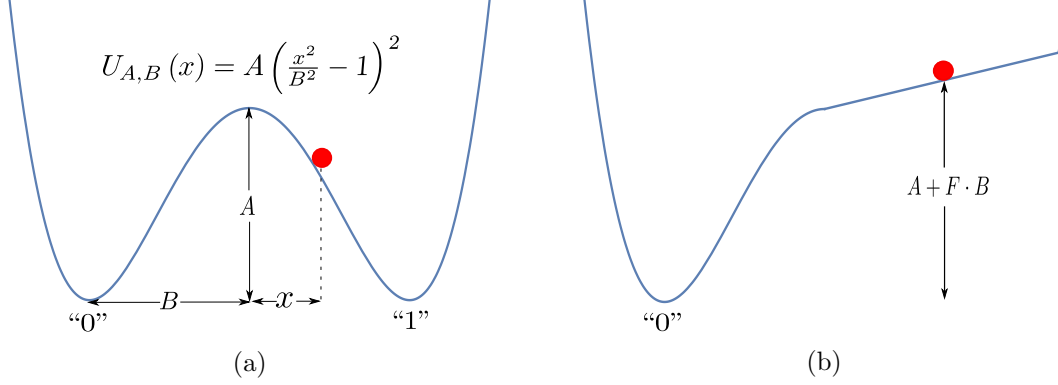
$$U_{A,B}(x) = A\left(\frac{x^2}{B^2} - 1\right)^2$$

"0" A B x "1"

(a)

$A + F \cdot B$

"0"

(b)

FIGURE 1: A bit as represented by a particle in a 1-D potential. Figure 1a: the bit in its resting state, with a barrier of height "A" separating particle locations that correspond to bit values of 0 or 1. Figure 1b: a control potential as in Example 2.1 is applied to erase the stored data.

2   The Double-Well Bit

We will represent a device to store one bit of information by a particle in a symmetric bistable potential $U_{A,B}(x) = A\left(\frac{x^2}{B^2} - 1\right)^2$, where A is the height of the well and $\pm B$ are the coordinates of the minima of the right and left wells. We will refer to the device as a whole by "a bit". The device reports "0" when the particle is in the left well, i.e., $x < 0$ and reports "1" otherwise (Figure 1a).

The dynamics of the particle is described by the Langevin equation.

$$\begin{aligned} m\,dx &= p\,dt \\ dp &= -\gamma p\,dt - \partial_x U_{A,B}(x)\,dt + \sqrt{2m\gamma k_B T}\,dW \end{aligned} \tag{1}$$

where m is the mass of the particle, x is position, p is momentum, γ is the friction coefficient of the medium, $U_{A,B}(x)$ is the potential, k_B is Boltzmann's constant, and T is the temperature of the heat bath. The term $\sqrt{2m\gamma k_B T}\,dW$ represents the effect of noise from the surroundings. The Langevin equation is a stochastic differential equation, to be mathematically interpreted as a Stratonovich integral. For our case both the Ito and Stratonovich interpretations coincide [24, pp. 109] since the noise coefficient $\sqrt{2m\gamma k_B T}$ does not depend upon p.

From [25, pp. 182], the generator for the Langevin equation 23 is

$$\mathcal{L} = \frac{p}{m}\partial_x - (\partial_x U_{A,B}(x))\,\partial_p + \gamma\left(-p\partial_p + k_B T \partial_p^2\right) \tag{2}$$

The **Hamiltonian** of the system is $H(x,p) = U_{A,B}(x) + \frac{p^2}{2m}$. The **Gibbs distribution**

$$\pi(x,p) = \frac{e^{-H(x,p)/k_B T}}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-H(x',p')/k_B T}\,dx'\,dp'} \tag{3}$$

is approached as the system relaxes to equilibrium. Convergence to $\pi(x,p)$ happens exponentially fast at a rate given by the first non-zero eigenvalue of the generator $\mathcal{L}$ [26].

3

2.1 Reliability

A device to store information should be able to store it with high fidelity for a specified long period of time. We introduce the **reliability time** to represent the time-scale over which our device can store data. Specifically, we define the reliability time τ_r as the expected first passage time for the particle to cross the barrier of the resting-state potential of the bit, given the Gibbs distribution $\pi(x, p)$ (Equation 3) as the initial distribution. That is,

$$\tau_r := \mathbb{E}\left[\inf\{t \geq 0 \mid x(t) = 0\}\right] \tag{4}$$

where the expectation is over trajectories $(x(t), p(t))$ distributed as specified by Equation 23 from the initial condition $(x(0), p(0)) \sim_{\text{law}} \pi(x, p)$. Note that τ_r is also the first passage time to cross the barrier for a bit prepared with a Gibbs distribution, but confined to either the left-hand well $\pi_0(x, p)$ or right-hand well $\pi_1(x, p)$.

$$\pi_0(x, p) = \begin{cases} 2\pi(x, p) & \text{if x} < 0 \\ 0 & \text{otherwise} \end{cases}, \quad \pi_1(x, p) = \begin{cases} 2\pi(x, p) & \text{if x} > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

Intuitively, once a typical particle has had enough time to reach the top of the barrier, the data stored is no longer reliable.

2.2 Setting information

A device intended to store information must provide functionality to load, or set this information into the device. Setting information is a two-bit operation. A common use case is when a reference bit and the bit to be set are initially at some arbitrary values. We require that after the SET operation the reference bit is unchanged whereas the bit to be set now holds a copy of the reference bit. This is the operation that Szilard [1] refers to as "copying" (in contrast, Landauer [2, 27] chooses to reserve the word "copying" for the operation where the bit to be set is initially already known to be in the state "0").

Note that in the operation of setting information, or copying in the sense of Szilard, initially the two bits are uncorrelated and unknown whereas after the operation they are still unknown but correlated. Thus implementing this operation requires decreasing the entropy of the system. Since it is easier to study a one-bit system rather than a two-bit system, we will investigate a one-bit proxy for the task of decreasing the entropy of the system, which is the task of **erasing** a bit.

Erasing involves taking a device whose initial state is maximally unknown into a known reference state, usually "0." Somewhat counter intuitively, given the name, erasing increases the information we know about the system. What is erased is not information but randomness. It helps to keep in mind the example of erasing a blackboard where some random state with chalk marks is reset to the "all clear" state.

2.2.1 Erasing

The example that follows describes a simple family of control potentials to implement the erasing operation for our device, which will form the basis of our analysis. One control potential from this family is illustrated in Figure 1b. We chose such a simple class of controls to make a full understanding feasible, setting a framework for analysing more complex protocols. We also note that arbitrary variation of a physical potential in reality is highly non-trivial;

experimental studies in which complex time-dependent potentials have been applied in fact use highly dissipate mechanisms to generate "effective" potentials [28, 29].

Example 2.1. Our control potentials are described by a single parameter $F \in \mathbb{R}_{>0}$ as follows.

$$V_F(x) := \begin{cases} A + F \cdot x - U_{A,B}(x) & \text{if } x \geq 0 \text{ and } A - U_{A,B}(x) + F \cdot x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

The Langevin equation in the presence of control is

$$m\,dx = p\,dt$$
$$dp = -\gamma p\,dt - \partial_x U_{A,B}(x)\,dt - \partial_x V_F(x)\,dt + \sqrt{2m\gamma k_B T}\,dW \tag{7}$$

Note that the control potential, as defined, is not differentiable at the boundary of the region in which it is non-zero. In practice, we assume that $\partial_x V_F$ changes rapidly but continuously in a small vicinity around these points.

In this work, we will consider variation of A, F and γ at fixed m, B, and T. In this case, m specifies the natural mass scale, B the natural length scale and $k_B T$ the natural energy scale; the natural time scale is then $\sqrt{mB^2/k_B T}$. Henceforth, all numerical quantities will be reported using reduced units defined with respect to these natural scales, although m, B and $k_B T$ will be retained within formulae.

2.2.2 Operational view of Erasing

The speed of bit operations is of practical importance: a useful bit must be reliable on much larger time-scales than those required to set or switch it. The control is switched on at time 0 and switched off at an appropriately-chosen time τ. The time τ is chosen beforehand, and does not depend on details of individual trajectories – a trajectory-dependent control would require measurement and feedback that itself would need accounting for [30, 31, 32, 33, 34, 35]. We could declare erasing as completed and switch off the control as soon as a majority of trajectories are expected to be in the left well. However, many of these "erased" bits would have high energies compared to typical bits drawn from the equilibrium distribution in the left well, $\pi_0(x, p)$. Thus they could rapidly return to the right well after a very short stay in the left well. So we insist on a more stringent condition. We require that the time τ should be large enough so that the majority of bits are in the target well, with an expected next passage time close to the reliability time.

One way to guarantee that the next passage time is high is by insisting on mixing, in the sense that the initial distribution $\pi(x, p)$ comes close to a distribution of particles thermalised in the left-hand well, $\pi_0(x, p)$. If this happens, we can guarantee that the expected next passage time will be equal to, or close to, the expected first passage time. However, we found this criterion too stringent for the following reason. At the end of the erasing protocol, it is not necessary that the distribution is close to $\pi_0(x, p)$ – only that the particles tend to relax to this distribution much faster than they cross back into the right-hand well, and thus have barrier passage times representative of particles initialised with $\pi_0(x, p)$. Nonetheless, we show in Section A.2.1 of the Supplementary Information that using such a criterion preserves the qualitative features reported below (in particular, the scaling of erasure time with friction in the high and low friction limits).

Instead, we define an **erasure region** in well "0" as all points (x, p) with total energy $H(x, p) \leq A - 3k_B T$ where A is the barrier height. We look for the average first passage time to reach the erasure region for particles initiated in well "1" and take this quantity to be representative of the erasing time-scale. The choice of $3k_B T$ criterion is somewhat arbitrary, but has been used before by Vega et al. [36] to study atom-surface diffusion. As we show in Section A.2.2 of the Supplementary Information, using $4k_B T$ makes no qualitative difference to our conclusions. This metric has the merit that it provides a clear computable criterion for erasing. Below, we demonstrate that particles within the $3k_B T$ erasure region do indeed have expected next passage times close to the reliability time, as required.

For a range of well parameters, we used the Langevin A algorithm from [37] [refer to Section A.1 of the Supplementary Information for integrator set-up and validation], to estimate $\tau(x, p)$, the average barrier crossing time for particles initialised at position x with momentum p in the left well, for a grid of points (x, p). The average reliability time for a given well can be approximated in terms of $\tau(x, p)$ as follows:

$$
\tau_r \approx \frac{\sum\limits_{x,p} \tau(x, p) \, e^{-H(x,p)/k_B T}}{\sum\limits_{x,p} e^{-H(x,p)/k_B T}}.
\tag{8}
$$

The deviation $\delta(x, p) := |1 - \frac{\tau(x,p)}{\tau_r}|$ for every point (x, p) in the grid is plotted in Figure 2, for a range of friction parameters at well height $A = 7$. It is clear that, for all values of friction, the points with total energy $H(x, p) \leq A - 3k_B T$ have reliability times close to τ_r. The same is true of other well heights A. This is because such particles typically undergo thermal mixing before they can escape the well. Once mixed, their next escape over the barrier will be on a time-scale of the order of τ_r.

Despite the robustness of this result to the value of the friction, the heatmaps in Figure 2 are friction-dependent. When γ is low, the particle diffuses very slowly in energy space, and it is the challenge of diffusing within this energy space that prohibits escape from the well. As a result the heatmap corresponding to $\gamma = 0.1$ (Figure 2a) follows the shape of constant energy contours. As friction starts increasing (e.g. in Figure 2b and 2c), diffusion in momentum-space becomes more rapid, but diffusion in position-space slows down. Once γ becomes very high (e.g. $\gamma = 100$ in Figure 2d), the behaviour of the heatmap is essentially determined by the initial position of the particle; those close to the barrier and with $U_{A,B}(x)$ sufficiently close to A can escape easily, but the momentum is irrelevant. Using the total energy $H(x, p)$ as a criterion ensures that we account for all the regimes of friction.

Since we are interested in the typical time scale of transferring particles to a different well from the existent well, we will sample initial points only from the right well. We define the **erasing time** τ_e as the expected time to hit the erasure region, given that the particle started in the right-hand well:

$$
\tau_e = \mathbb{E}\left[\inf\{t \geq 0 \mid x(t) < 0 \text{ and } H(x(t), p(t)) \leq A - 3k_B T\}\right]
\tag{9}
$$

where $(x(t), p(t))$ is the solution to Equation 7 with the initial condition $(x(0), p(0)) \sim_{\text{law}} \pi_1(x, p)$. Given this definition, τ_e indicates a typical time scale over which the control must be applied to successfully erase a large fraction of the bits. In practice, the control would be applied for a period $\tau > \tau_e$ to achieve high accuracy. We will use τ_e as an indicative time scale

(a) $\delta(x,p)$ when $\gamma = 0.1$

(b) $\delta(x,p)$ when $\gamma = 1$

(c) $\delta(x,p)$ when $\gamma = 10$
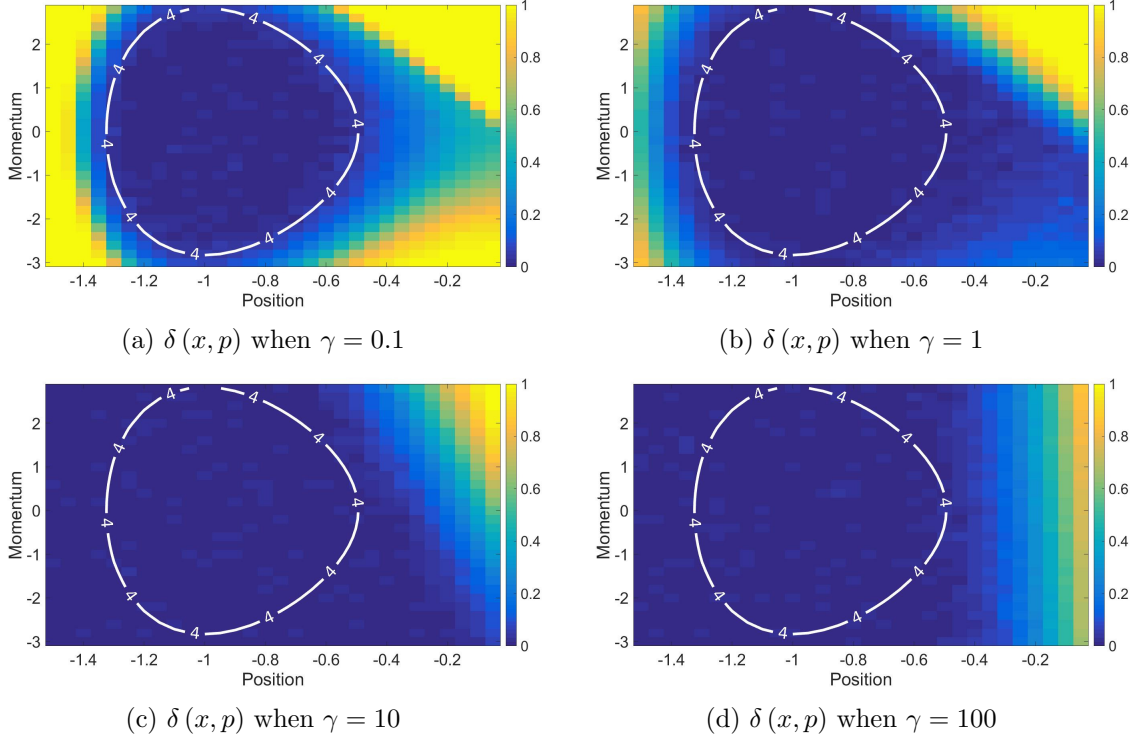
(d) $\delta(x,p)$ when $\gamma = 100$

FIGURE 2: For particles initiated with $H(x,p) \leq A - 3k_BT$, well escape times are close to τ_r. Heat maps show the fractional deviation in expected escape time $\delta(x,p)$ from the well-thermalised average τ_r, as a function of initial position x and momentum p. The labelled contours correspond to a well height $A = 7$ with energy $H(x,p) = A - 3k_BT = 4k_BT$. These heat maps are representative of the situation for other barrier heights $A \geq 5k_BT$.

of control operation for the purposes of our analysis. It is useful to decompose the erasing time τ_e as the sum of two times: the **transport time** and **mixing time**.

- **Transport time (τ_t):** The time taken by the particle to reach well "0" given that it is initially distributed according to $\pi_1(x,p)$.

$$\tau_t = \mathbb{E}\left[\inf\{t \geq 0 \mid x(t) \leq 0\}\right] \tag{10}$$

where $x(t)$ is the solution to Equation 7 with the initial condition $(x(0), p(0)) \sim_{\text{law}} \pi_1(x,p)$.

- **Mixing time (τ_m):** The time taken by the particle to mix sufficiently inside the well. This is the time starting from when the particle first reaches well "0" to when it first hits the erasure region.

$$\tau_m = \tau_e - \tau_t \tag{11}$$

2.2.3 Cost of erasing

In this section, we calculate the work done in erasing a bit. From Sekimoto's expression [38, 39], for a protocol applied for a time τ and with a region of effect $I = \{x \geq 0 \mid A - U_{A,B}(x) +$

$F \cdot x \geq 0\}$,

$$\langle W \rangle := \int\limits_{0}^{\tau} \int\limits_{x \in I} \frac{\partial V_F(x,t)}{\partial t} p(x,t) \, dx \, dt, \tag{12}$$

where $p(x,t)dxdt$ is the probability that the particle is between position $(x, x+dx)$ in the time interval $(t, t+dt)$. There are two potential sources of work that appear in our calculation.

1. When we begin the erasure protocol by switching on the control to lift the particle.

2. At the end of the protocol when we switch off the control.

We note that in our family of controls, there is negligible energy recovered when the control is switched off (refer to Section A.4 of the Supplementary Information), since the probability of the particle being in the region in which the control is applied is small. More generally, the question of whether energy might be recovered from small systems and stored efficiently is a complex one, despite the optimism shown in previous discussions of erasing. Indeed, current technology does not attempt to recover any energy from bits.

We now calculate the work done for our protocol, Example 2.1. The particle's initial potential energy is approximately $k_B T/2$ on average, due to the equipartition theorem, and after the control is switched the average potential energy is $A + F \cdot B$ for a particle in the right well, since the particle is localised around $x = B$, and still $k_B T/2$ for a particle in the left well. So, ignoring energy recovery at the end of the operation, the net work done for the erasure protocol is $W = (A + F \cdot B - k_B T/2)/2$. As justified analytically and numerically in Section A.3.1 of the Supplementary Information, this approximation is accurate for the values of A and F that we consider, and we will use this as the form of work for the rest of the manuscript.

Observation 2.2. Work is an increasing function of well height A at fixed F and γ. This follows immediately from the expression of work $W = (A + F \cdot B - k_B T/2)/2$.

3 Friction-based trade-offs for reliability and erasing

We explore the behaviour of the reliability and erasing time-scales as functions of the friction coefficient. We find that both these time-scales are non-monotonic, roughly U-shaped functions of the friction coefficient. A high reliability time requirement is favoured by a very low or very high friction; whereas a low erasing time requirement is helped by the choice of a moderate value of friction. Since a bit designer would seek reliable bits (needing high or low friction) that can be erased fast (needing intermediate friction), this yields a friction-based trade-off between reliability and speed of erasure.

3.1 Reliability Time

Our definition of reliability time (Equation 4) is very similar to the classic problem of escape rates from one-dimensional wells (Fig 3), as applied in transition state theory to understand chemical reactions. In a famous paper [40], Kramer found analytic expressions for the escape rate k from a well by calculating the flux of particles between a source on one side of the

barrier (x_A) and a sink at the other side (x_B). Kramer's expressions apply separately to the regimes of low friction, moderate to high friction and very high friction. Later the groups of Melnikov and Meshkov [41] and Pollack, Grabert and Hänggi [42] gave formulae that interpolate accurately over all values of friction (see review in [43]). We will apply the result of Melnikov and Meshkov to estimate analytical forms of the escape rate for our bistable system

$$k = \frac{\omega_0}{2\pi} \left[\sqrt{1 + \frac{\gamma^2}{4\omega_b^2}} - \frac{\gamma}{2\omega_b} \right] g \, e^{-A/k_B T}, \text{where}$$

$$\ln g = \frac{1}{2\pi} \int_0^{\frac{\pi}{2}} \ln \left[1 - \exp \left(\frac{-\gamma \, I(A)}{4 k_B T \cos^2 x} \right) \right] dx. \tag{13}$$

Here, ω_b is the angular frequency at barrier height, ω_0 is the angular frequency at the bottom of the well and $I(A)$ is the action for barrier height A. Refer to Section A.6 in the Supplementary Information for a detailed definition of these parameters and calculations for our system.
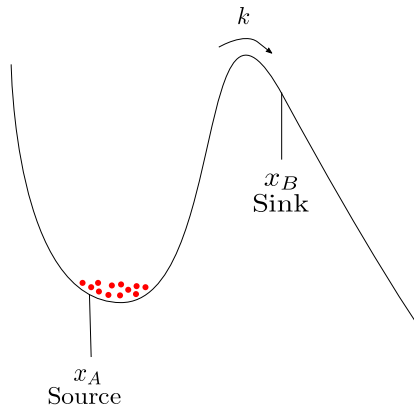

FIGURE 3: The escape of particles from a one-dimensional well. Kramer [40] considered a source of particles at the bottom of the well, and estimated the rate of escape to a sink on the far side of a barrier.

We plot the analytical prediction of $1/k$ given by Eq. 13 in Fig. 4 for two values of well height A, as a function of friction γ. This prediction is compared to average first passage time for particles to reach the top of the barrier from an initial Boltzmann distribution within a single well. The two quantities differ at large γ because Kramer's definition does not treat a particle that crosses the barrier but then immediately crosses back as having "escaped", whereas our definition of reliability in terms of a first passage time treats such particles as no longer being reliable. In the underdamped regime, immediate recrossings are rare and hence τ_r and $1/k$ coincide; in the overdamped regime, particles that reach the barrier top have a 50% chance of returning and so $\tau_r = 1/2k$. As can be seen from Fig. 4, τ_r smoothly interpolates between $\frac{1}{k}$ and $\frac{1}{2k}$, with the small numerical factor providing only a minor correction to the underlying physics of the analytical expression in Eq. 13.

The Melnikov-Meshkov expression predicts an almost-exponential scaling of $1/k$ with barrier height A, which is reproduced by τ_r and expected from the Arrhenius rate law [44]. Note
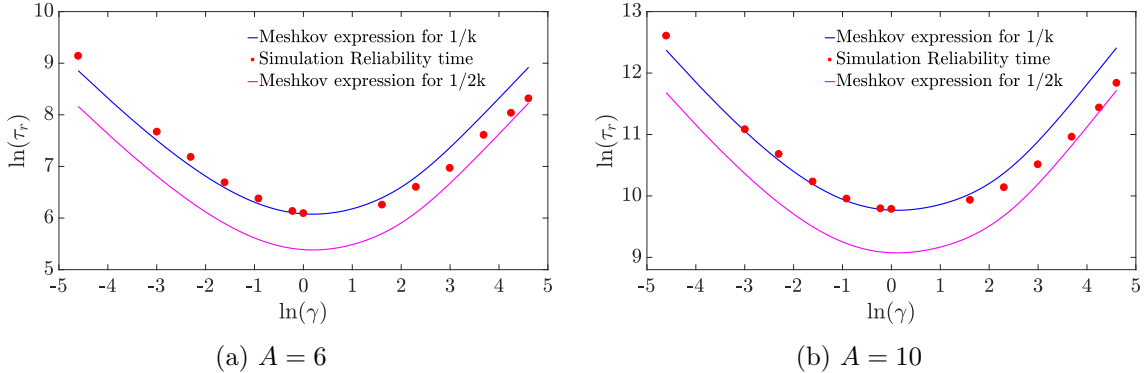
(a) $A = 6$ (b) $A = 10$

FIGURE 4: The reliability time $\tau_r \propto \frac{1}{\gamma}$ in the low friction regime, $\tau_r \propto \gamma$ in the high friction regime and is minimum at moderate friction. Simulation results are compared to the inverse of escape rate from a single well $(1/k)$ and $(1/2k)$, as predicted by Eq. 13. Here, and elsewhere in the manuscript, error bars are omitted when comparable to data points.

that both $1/k$ and τ_r are non-monotonic in friction γ, with long reliability times in the underdamped and overdamped limits. This behaviour results from the need for particles to diffuse in both position and energy in order to reach the top of the barrier from an initial state thermalized within a single well. At high friction, particles rapidly sample different kinetic energies due to strong coupling with the environment, but move slowly in position space and hence take a long time to cross the barrier. At low friction, particles can move rapidly but their energy remains effectively constant over short time periods. They only cross the barrier when they have eventually gained enough total energy. Intermediate friction, when neither process is excessively slow, gives the shortest τ_r. This behaviour is typical of equilibrating systems in which an initial out-of-equilibrium condition (particles are guaranteed to be on one side of the well and not the other) relaxes towards an equilibrium state (particles on both sides of the barrier), and is thus insensitive to the details of our bit design.

A more detailed analysis of the dependence of the reliability time on various parameters, and indeed the functional form of the well, is possible. However, these details are not necessary for the conclusions we draw in the rest of this manuscript, and hence we omit them here.

3.2 Erasing time

As noted earlier, the erasing time is composed of two parts: the transport time defined in Eq. 10 and the mixing time defined in Eq. 11. We now present analytical estimates of these times and compare them with numerical solutions.

3.2.1 Transport time

We can obtain an analytical estimate of the transport time in low and high friction limits by assuming that a particle starting at $x = B$ moves deterministically under the influence of the potential slope and drag force.

1. **Low friction regime**: The particle travels with a constant acceleration of $\frac{F}{m}$ and the time taken to travel a distance B is $\tau_t \approx \sqrt{2mB/F}$.

10

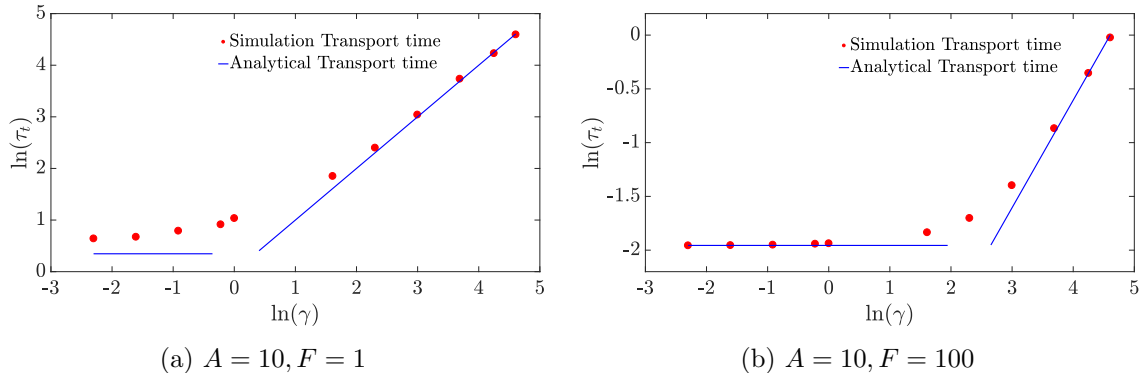(a) $A = 10, F = 1$ (b) $A = 10, F = 100$

FIGURE 5: The transport time obtained from simulations approximates the analytical estimates of $\tau_t \approx \sqrt{2mB/F}$ in the low friction regime, $\tau_t \approx mB\gamma/F$ ($\propto \gamma$) in the high friction regime.

2. **High friction regime**: In this regime, we assume that the net force on the particle (arising from the sum of drag and potential) is zero. The particle travels with a velocity of $F/m\gamma$, and the time taken to travel a distance B is $\tau_t \approx mB\gamma/F$.

We thus expect the transport time to be constant in the underdamped regime and increase linearly with friction in the overdamped regime. Figure 5 illustrates that this scaling is observed in Langevin simulations, and that numerical values are in reasonable agreement with these crude estimates. The largest quantitative deviations occur at low force and low friction (e.g. $F = 1$ in Figure 5a), when the diffusion of the particle on the slope contributes significantly to τ_t. This results in a simulation transport time larger than the analytical estimate.

3.2.2 Mixing time

Similar to the transport time, analytical estimates of the mixing time can be obtained in the limits of high and low friction.

1. **Low friction regime**: For purposes of approximate calculation we treat the well "0" as a harmonic oscillator. Deterministically, the energy of a harmonic oscillator decays exponentially in the underdamped regime. Therefore we have $E(t) = E_0 \, \mathrm{e}^{-\gamma t}$, where E_0 is the initial energy of the particle when it first reaches $x = 0$ and $E(t)$ is the energy of the particle at time t. In the underdamped regime, a particle starting at B arrives at position $x = 0$ with energy $E_0 \approx A + F \cdot B$. Thus solving for $E(\tau_{\mathrm{mix}}) = A - 3k_BT$,

$$\tau_{mix} \approx \frac{1}{\gamma} \log \frac{A + F \cdot B}{A - 3k_BT} \tag{14}$$

2. **High friction regime**: A sensible estimate of the behaviour can be obtained by explicitly modelling the diffusion of the particle near the barrier top. In the overdamped limit, the criterion of reaching a total energy of $E(\tau_{\mathrm{mix}}) = A - 3k_BT$ is equivalent to reaching a point d which has potential energy of $A - 3k_BT$, since momenta are sampled arbitrarily rapidly in this limit. To proceed, we consider the typical time required to
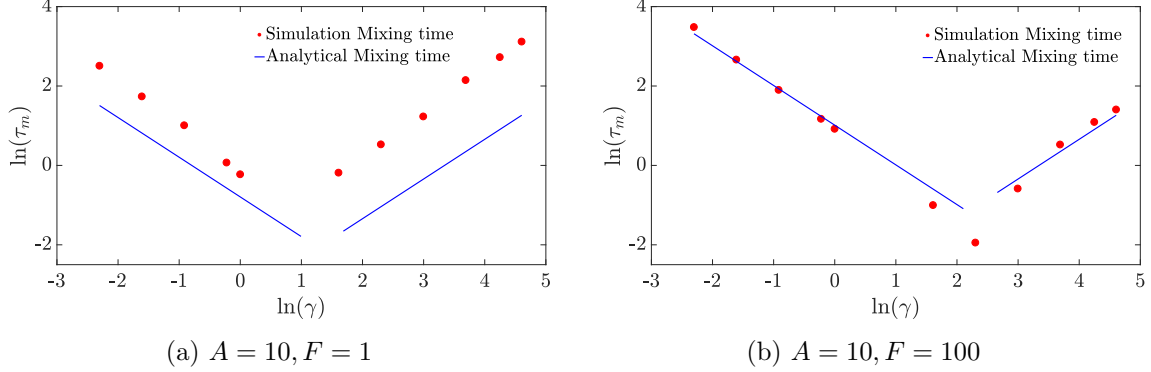
11

(a) $A = 10, F = 1$ (b) $A = 10, F = 100$

FIGURE 6: Evidence from simulation that the mixing time $\tau_{mix} \approx \frac{1}{\gamma} \log \frac{A+F \cdot B}{A-3k_BT} \left(\propto \frac{1}{\gamma} \right)$ in the low friction regime, $\tau_{mix} \approx \frac{mB^2\gamma}{2\sqrt{2A}} (\propto \gamma)$ in the high friction regime and is minimised at moderate friction.

reach an absorbing barrier at d starting from $x = 0$, assuming a sufficiently large F that we can treat $x = 0$ as a reflecting barrier. Starting from the overdamped stochastic differential equation

$$m\gamma \, dx = -\partial_x U_{A,B}(x) \, dt + \sqrt{2m\gamma k_B T} \, dW \tag{15}$$

with generator $\mathcal{L} = \frac{k_B T}{m\gamma} e^{\frac{U_{A,B}(x)}{k_B T}} \partial_x e^{\frac{-U_{A,B}(x)}{k_B T}} \partial_x$, we apply the standard methods outlined in Pavliotis [25, (7.1), pp. 239], which leads to the following system of equations for the average mixing time $\tau_{mix}(x)$ as a function of the initial position x

$$\frac{k_B T}{m\gamma} e^{\frac{U_{A,B}(x)}{k_B T}} \partial_x e^{\frac{-U_{A,B}(x)}{k_B T}} \partial_x \tau_{mix}(x) = -1, d < x \leq 0. \tag{16}$$

$$\tau_{mix}(x) = 0, x = d.$$

We can solve Equation 16 using appropriate limits to get

$$\tau_{mix}(x) = \frac{m\gamma}{k_B T} \int_0^{\sqrt{\frac{3k_B T}{2A}} B} \int_0^q e^{\frac{U(q)-U(r)}{k_B T}} \, dq dr, \tag{17}$$

where we have approximated the potential near the barrier as an inverted harmonic oscillator to estimate $d = \sqrt{\frac{3k_B T}{2A}} B$. Repeating this approximation within the integral, we obtain

$$\tau_{mix}(x) \approx \frac{m\gamma}{k_B T} \int_0^{B\sqrt{\frac{3k_B T}{2A}}} \int_0^q e^{\frac{2A(r^2-q^2)}{B^2 k_B T}} \, dq dr \approx \frac{mB^2\gamma}{2\sqrt{2A}}. \tag{18}$$

Equations 14 and 18 predict that the mixing time will scale as $1/\gamma$ in the low friction limit and as γ in the high friction limit. In the first case, mixing within the well is limited by the rate at which the particle can reduce its total energy, whereas in the second it is determined by the speed with which the particle can diffuse in position space to a configuration

with lower potential energy. We plot simulation results for the mixing time, along with the analytic predictions, in Fig. 6, confirming this scaling and the resultant non-monotonicity. Quantitatively, simulation results deviate from the crude analytic predictions at low force (e.g. $F = 1$ in Figure 6a), when it is no longer reasonable to treat $x = 0$ as either a reflecting barrier or a steep side of a harmonic well. Instead, excursions of the particle back onto the slope occupying the region $x > 0$ lead to much larger mixing times. Nonetheless, the scaling and non-monotonicity in friction are preserved. Combining τ_{trans} and τ_{mix} gives τ_e, plotted in Fig. 7. Analytically, the erasing time is given as:

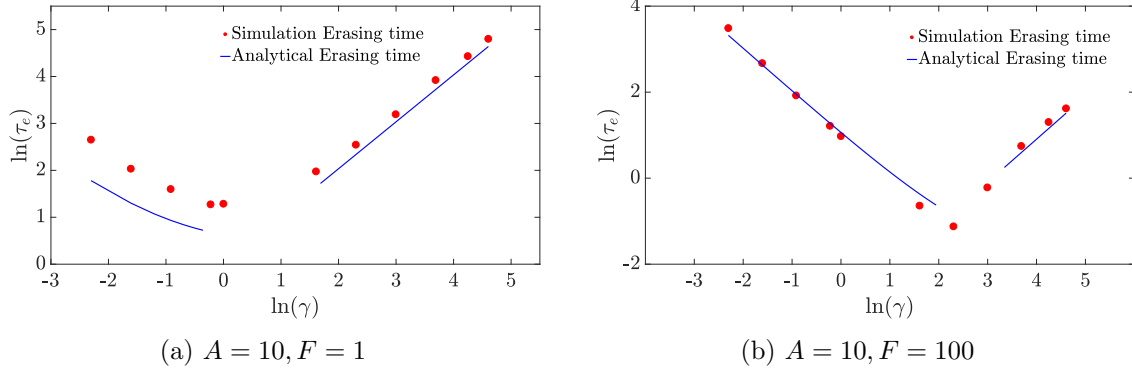(a) $A = 10, F = 1$ (b) $A = 10, F = 100$

FIGURE 7: Evidence from simulation that the erasing time $\tau_e \approx \sqrt{\frac{2mB}{F}} + \frac{1}{\gamma} \log \frac{A+F \cdot B}{A - 3k_B T}$ in the low friction regime, scaling as $1/\gamma$, and $\tau_e \approx \frac{mB\gamma}{F} + \frac{mB^2\gamma}{2\sqrt{2}A}$ in the high friction regime, scaling as γ. The erasing time is minimised at moderate friction.

1. **Low friction regime**:

$$\tau_e \approx \sqrt{\frac{2mB}{F}} + \frac{1}{\gamma} \log \frac{A + F \cdot B}{A - 3k_B T}. \tag{19}$$

2. **High friction regime**:

$$\tau_e \approx \frac{mB\gamma}{F} + \frac{mB^2\gamma}{2\sqrt{2}A} \tag{20}$$

Like reliability, erasing time is large in the underdamped and overdamped limits, and minimized at intermediate values of friction. The physical cause is the same as before; our erasing protocol involves setting the system into a non-equilibrium state, and waiting for the system to relax towards an equilibrium in the perturbed potential. This process requires the system to diffuse in energy space and also explore configuration space, and is therefore favoured by intermediate friction. Specifically, if the friction is too low, the particle oscillates and slowly loses energy to be confined within the desired well. If the friction is too high, both the transport and mixing times increase as the particle's movement through space is so slow. The relative importance of these effects can be seen in Fig. 8. We note that the value of the damping γ that minimises τ_e is quite sensitive to F (Fig. 8). Fundamentally, a larger F means the challenge of moving in position-space is made easier, and a greater loss of energy is needed to reach equilibrium. Therefore a higher friction coefficient is optimal. As with the

reliability time, further analysis is possible but not necessary for the conclusions we wish to draw. Once again, the key point is the trade-off between high and low friction, which is not specific to our control. Indeed, is likely to be quite generic since any protocol will necessary push the system out of equilibrium, and will require particles to be typically confined within the target well before the control is removed.

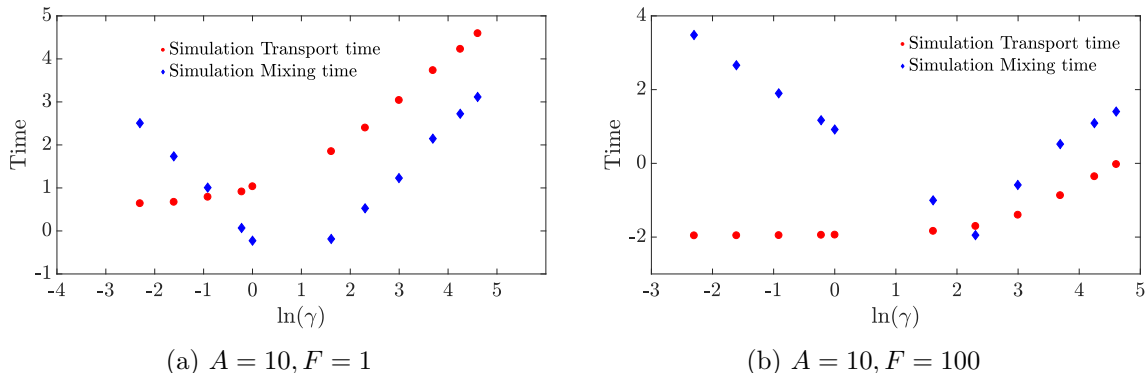(a) $A = 10, F = 1$ (b) $A = 10, F = 100$

FIGURE 8: Comparison of transport and mixing times. Transport time dominates the mixing time for low force at high friction.

Both erasing and reliability times exhibit a trade-off in friction, being minimised by intermediate values. This fact sets up a second trade-off between designing bits with extreme values of friction to optimise reliability, or moderate values of friction to optimise erasing. The consequences of this secondary trade-off will be explored in Section 4.

3.2.3   Additional dependencies of the erasing time

A larger value of A implies a steeper descent into the target left-hand well, making mixing faster. We therefore expect that the mixing time and hence the erasing time monotonically decreases with A.

Observation 3.1. The erasing time is a strictly decreasing function of well height A at fixed F, γ. This can be seen from the analytic expressions of erasing time (Equations 19 and 20) backed up with numerical simulations (Figure 9).

By contrast, erasing time shows a non-monotonic dependence on F at fixed A, γ. Applying too little force leads to slow transport, and doesn't effectively trap the particle within the target well. But applying too much force supplies the particle with too much energy, which must subsequently be lost during the mixing period. The fact that erasing time monotonically decreases with A at fixed F and γ, and shows a non-monotonic dependence on F at fixed A and γ, leads to non-monotonic dependence of τ_e on F at fixed $W = A + F$ and γ. We illustrate this non-monotonicity in Figure 10, in which simple regression formulae have been fitted to the simulation data to enable interpolation at fixed W and γ (see Section A.5 of the Supplementary Information). As friction increases, the force required to provide the particle with excess energy increases, leading to minima at higher values of F.

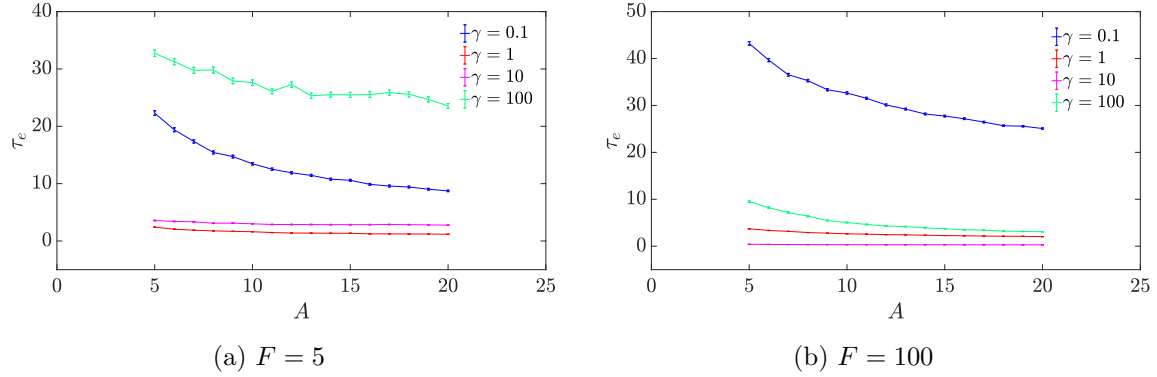We make the following observation which will be used in the subsequent section.

(a) $F = 5$

(b) $F = 100$

FIGURE 9: Evidence that the erasing time is a strictly decreasing function of well height across a range of F and γ. Other values of F and γ show similar behaviour.

(a) $\gamma = 0.1$

(b) $\gamma = 1$

(c) $\gamma = 10$
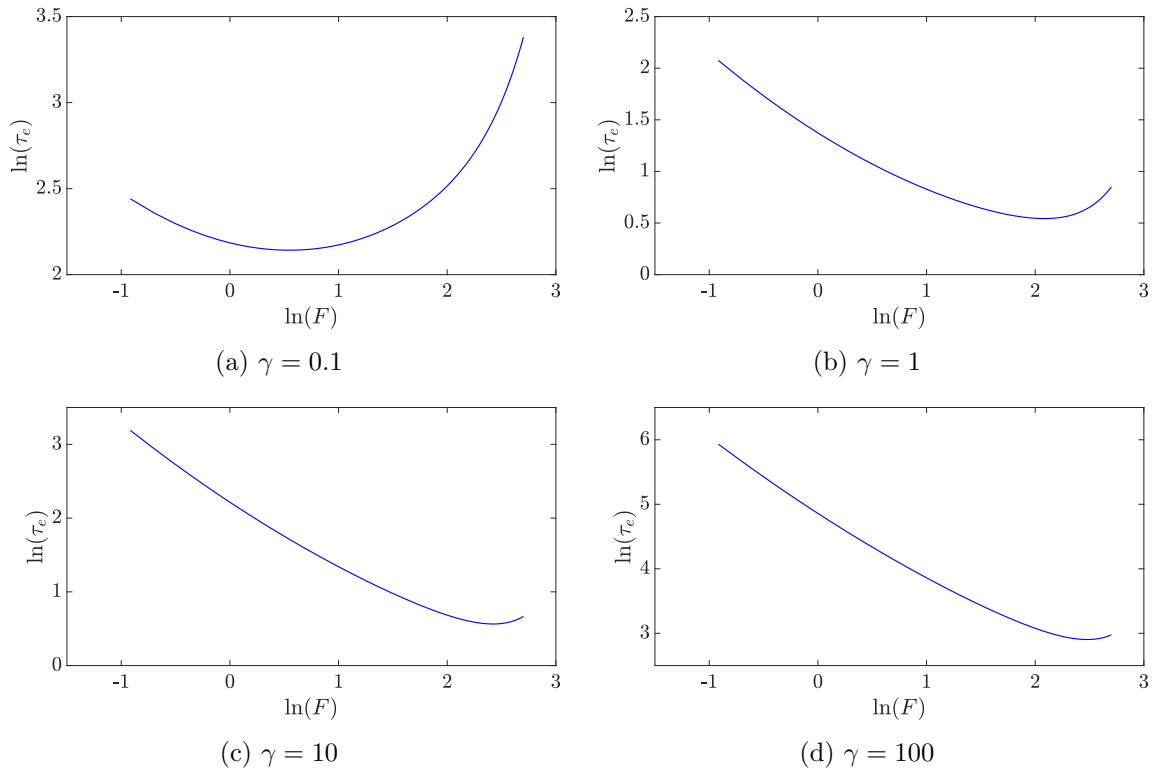
(d) $\gamma = 100$

FIGURE 10: For a fixed value of W and γ, the erasing time is a non-monotonic function of F and is minimum at moderate F. This is illustrated at work $W = 20$ for various values of γ.

Observation 3.2. We have found no evidence of multiple local minima of erasing time in a level set of work for our control family. Refer to Section A.7 of the Supplementary Information for characterstic plots showing the minima of erasing time in a level set of work. Physically this is unsurprising since the non-monotonicity in τ_e with γ and F mentioned above arise from fairly simple trade-offs, producing curves with single minima.

As with the reliability time, a more detailed analysis of the dependence of τ_e on other parameters, and even the shape of the control, is possible. However, these details are likely to be difficult to generalise, and are not necessary for the conclusions we draw in the subsequent sections.

4   Design of Bits

We are now ready to study the question of how to design good bits. A design involves choosing parameters A, F, γ for a bit to satisfy requirement specifications in terms of speed of erasing and reliability, without expending more work than required. The most general formulation of our problem would require us to also allow the length scale B, the temperature T and the mass m to vary, as well as allowing arbitrary controls. Such a formulation would appear to make the problem even more challenging, so it seems prudent in a first analysis to restrict our analysis to the variables A, F, and γ. Our restricted analysis is not without value since the underlying technology in any given construction typically does not allow arbitrary variation. Our numerical analysis with Example 2.1 will guide us in our assumptions and analysis, but our results will hold in greater generality. We will construct our proofs based on general assumptions, and subsequently explain how these assumptions are met by our control family.

We introduce the following terms.

1. The design of a bit is completely specified by the **design triple** (A, F, γ). **Design Space** ($\mathcal{DS}$) is the space of all design triples (A, F, γ).

2. A **requirement specification** is a tuple $(t_r, t_e) \in \mathbb{R}^2_{>0}$ denoting the reliability and erasing time that we require of the bit. **Requirements Space** ($\mathcal{RS}$) is the space of all requirement specifications.

3. **Erasing time** $\tau_e : \mathcal{DS} \to \mathbb{R}_{>0}$ takes a design triple (A, F, γ) to the time required for erasing the corresponding bit under the control protocol specified by F. **Reliability time** $\tau_r : \mathcal{DS} \to \mathbb{R}_{>0}$ takes a design triple (A, F, γ) to the reliability time of the corresponding bit. Note that τ_r is constant as a function of F since it is a property of the dynamics in the absence of control.

4. **Work** $W : \mathcal{DS} \to \mathbb{R}_{>0}$ represents the expected work done by the control in erasing the corresponding bit. We will assume that W is constant as a function of γ, as is the case in Example 2.1.

5. A design (A, F, γ) is **feasible** for a requirement (t_r, t_e) iff both $\tau_r(A, F, \gamma) \geq t_r$ and $\tau_e(A, F, \gamma) \leq t_e$. A (t_r, t_e)-feasible design (A, F, γ) is (t_r, t_e)-**optimal** iff the work $W(A, F, \gamma)$ is minimum among all (t_r, t_e)-feasible designs.

6. Inspired by the observation that non-trivial minima of erasing time at fixed work exist for our family of protocols (Section 3.2.3), we define the notion of **trapped** bits. A design (A, F, γ) is **trapped** iff for all designs (A', F', γ') with $W(A, F, \gamma) = W(A', F', \gamma')$, the erasing time $\tau_e(A, F, \gamma) \leq \tau_e(A', F', \gamma')$. A design (A, F, γ) is **uniquely trapped** iff for all designs (A', F', γ') with $W(A, F, \gamma) = W(A', F', \gamma')$, the erasing time $\tau_e(A, F, \gamma) \leq \tau_e(A', F', \gamma')$ with equality iff $(A, F, \gamma) = (A', F', \gamma')$. A design (A, F, γ) is **locally trapped** iff there exists a neighbourhood of (A, F, γ) consisting of bits (A', F', γ') with $W(A, F, \gamma) = W(A', F', \gamma')$ such that the erasing time $\tau_e(A, F, \gamma) \leq \tau_e(A', F', \gamma')$. More

informally, a trapped design has the lowest erasing time within a level set of work; a trapped design is unique if it is the *only* design within that level set of work to have the minimal erasing time; and a locally trapped design has the minimal erasing time within a local neighbourhood of designs of equal work.

7. A requirement specification (t_r, t_e) is **unsaturated** iff there exists a (t_r, t_e)-optimal design (A, F, γ) such that either $\tau_r(A, F, \gamma) > t_r$ or $\tau_e(A, F, \gamma) < t_e$. A feasible requirement specification that is not unsaturated is called **saturated**.

Throughout this section, we will assume that τ_e, τ_r, and W are continuous functions.

We will state the main results related to the properties of the optimal design leaving the detailed proofs to the Supplementary Information. We first claim that an optimal design always saturates the bound on the erasing time constraint. Further, if the optimal bit is not locally trapped, then it also saturates the bound on the reliability time constraint.

Claim 4.1 (Saturation of timescales). Let us assume that it is possible to locally decrease work at fixed reliability time (This is generally possible since one can perturb the control parameters to reduce work; but reliability time does not depend on the control parameters). Fix requirement specifications $(t_r, t_e) \in \mathcal{RS}$. Suppose (A, F, γ) is a (t_r, t_e)-optimal design. Then

1. $\tau_e(A, F, \gamma) = t_e$.

2. If the design (A, F, γ) is not locally trapped, then $\tau_r(A, F, \gamma) = t_r$.

Proof. Refer Section A.8 in the Supplementary Information. □

The next claim provides insight into the geometry of optimal designs. In particular, it states that under mild assumptions the requirement space is divided into two regions by a boundary given by the reliability and erasing times of trapped designs. Requirements with $t_r < t'_r$ and $t_e = t'_e$, where (t'_r, t'_e) is a requirement on the dividing line, are unsaturated, while other requirement specifications are saturated.

Claim 4.2 (Saturated and Unsaturated Requirements). Assume that the erasing time of trapped designs is a strictly decreasing function of the work (refer to Observation A.1 in the Supplementary Information for a justification), and that as before it is always possible to decrease work at fixed reliability time. Let (A^*, F^*, γ^*) be a trapped design such that $\tau_e(A^*, F^*, \gamma^*) = t_e$.

1. If $t_r \leq \tau_r(A^*, F^*, \gamma^*)$ then (A^*, F^*, γ^*) is (t_r, t_e)-optimal.

2. If $t_r < \tau_r(A^*, F^*, \gamma^*)$ then (t_r, t_e) is unsaturated.

3. Make the additional assumption that locally trapped designs are uniquely trapped (as noted for our family of protocols (Example 2.1) in Observation 3.2).

   If $t_r \geq \tau_r(A^*, F^*, \gamma^*)$, then (t_r, t_e) is saturated.

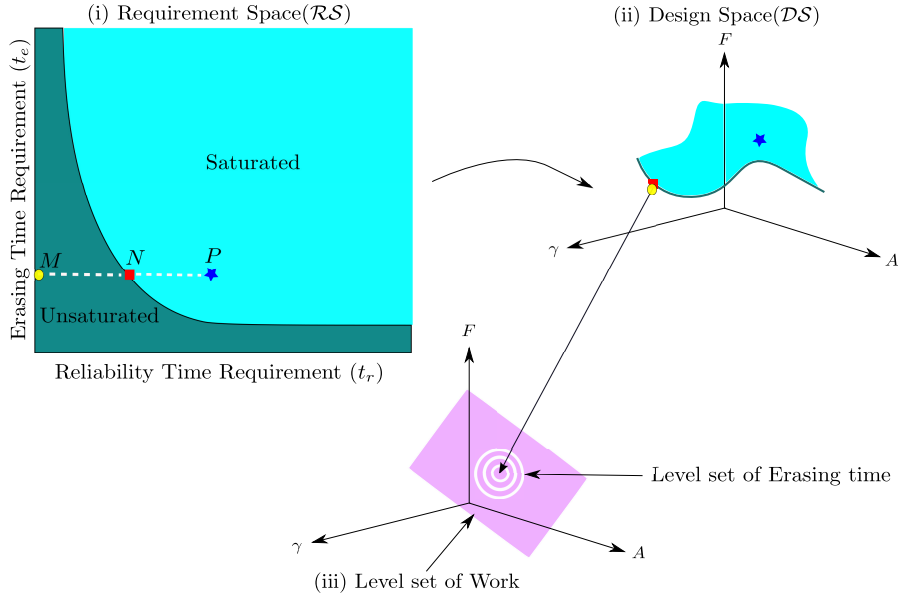*Proof.* Refer Section A.8 in the Supplementary Information. □

17

FIGURE 11: An illustration of the mapping of requirement specifications to optimal designs. The design space is divided by a curve corresponding to the reliability and erasing times of trapped designs. Points M, N in (i) Requirement Space($\mathcal{RS}$) having the same erasing time requirement get mapped to the same optimal bit in (ii) Design Space($\mathcal{DS}$): a trapped design with τ_r and τ_e equal to the requirements at N. The requirement specifications represented by points like P having the same t_e but greater t_r than N are mapped to distinct points in design space. (iii) A representation of a level set of work W within $\mathcal{DS}$, illustrating that the optimal designs to which unsaturated requirements are mapped minimize erasing time among all designs requiring the same work.

The claims about saturation/unsaturation of times-scales can also be proved using KKT conditions [Refer Section A.9 in the Supplementary Information], a standard tool from optimization theory.

A more intuitive picture of the results can be understood from Figure 11. In this figure, we illustrate how finding an optimal design subject to a specification maps a point in the requirement space to a point in the design space. For a trapped design (A^*, F^*, γ^*), requirements with $t_r < \tau_r(A^*, F^*, \gamma^*)$ and $t_e = \tau_e(A^*, F^*, \gamma^*)$ are unsaturated and get mapped to the same design (A^*, F^*, γ^*) (claims 4.2. 2 and 4.2. 1). If the design (A^*, F^*, γ^*) is uniquely trapped, then requirements with $t_r \geq \tau_r(A^*, F^*, \gamma^*)$ and $t_e = \tau_e(A^*, F^*, \gamma^*)$ are saturated (claim 4.2. 3).

Figure 12 illustrates these results for our example family of controls (Example 2.1). As discussed in Section A.5 of the Supplementary Information, we have implemented simple regression to fit the functions $\tau_e(.)$ and $\tau_r(.)$ to our simulation results. We then identified trapped designs using numerical minimisation, plotting the requirement specifications saturated by these designs. For each trapped bit (A^*, F^*, γ^*), we randomly selected requirements with $t_e = \tau_e(A^*, F^*, \gamma^*)$, but with t_r either greater than equal to or less than $\tau_r(A^*, F^*, \gamma^*)$, and used numerical optimization techniques to search for the optimal designs. The results
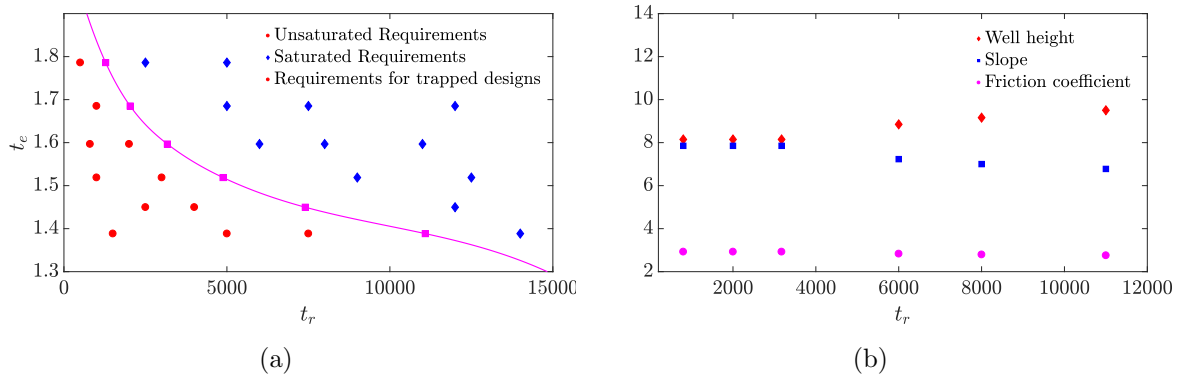
FIGURE 12: Illustration of the division of Requirement Space($\mathcal{RS}$) into saturated and unsaturated regions by requirements that correspond to trapped designs. (a) Squares show requirements (t_e, t_r) that are saturated by trapped designs for the family of protocols we consider. Numerical optimization shows that requirements to the left of the locus defined by these points are unsaturated (circles), whereas requirements to the right are saturated (diamonds). (b) A plot of the optimal designs for points from (a) at $t_e = 1.5967$. It is clear that for requirements $t_r \leq 3173$, lying to the left of the trapped-design locus in (a), optimal design parameters are identical whereas they are distinct for $t_r > 3173$.

support our analysis; requirements with $t_r < \tau_r(A^*, F^*, \gamma^*)$ are unsaturated, and those with $t_r \geq \tau_r(A^*, F^*, \gamma^*)$ are saturated. Furthermore, as we show in Figure 12 (b), unsaturated requirements at fixed t_e all map to the same trapped design.

4.1 Optimal friction for simple controls

In Section 3, we demonstrated that both reliability and erasing times are non-monotonic in friction, with short erasing times favoured by moderate values of friction, and long reliability times favoured by extreme values. In what follows, we give a precise quantification of the resultant trade-off in finding the friction of an optimal bit. The analysis is significantly simplified for our family of controls, in which work is independent of the friction coefficient.

Let us introduce the following terms. Fix an A and F. Then,

1. γ_{crit}^e is the friction coefficient that minimizes erasing time as a function of friction coefficient γ at fixed A and F, i.e., for all $\gamma' \in \mathbb{R}_{>0}$, we have:

$$\tau_e(A, F, \gamma_{\text{crit}}^e) \leq \tau_e(A, F, \gamma'). \tag{21}$$

We call the design $(A, F, \gamma_{\text{crit}}^e)$ **critically damped**.

2. γ_{crit}^r is the friction coefficient that minimizes reliability time as a function of friction coefficient γ at fixed A and F, i.e., for all $\gamma' \in \mathbb{R}_{>0}$, we have:

$$\tau_r(A, F, \gamma_{\text{crit}}^r) \leq \tau_r(A, F, \gamma'). \tag{22}$$

It is easy to note that trapped bits are also critically damped. In Figure 13 we show illustrative curves of the erasing and reliability times as a function of friction coefficient γ at

19

fixed A, F. These curves have single minima at γ^e_{crit} and γ^r_{crit}, respectively. Also shown on this graphs are regions of friction space that can be eliminated from consideration for optimal bits. To eliminate extreme values of friction, we note that the design must have a minimal finite A to be a well-defined two-state system in the resting state. For our bit, it is $A_{\min} \approx 3$. In the next claim we precisely describe which regions of friction can be eliminated.
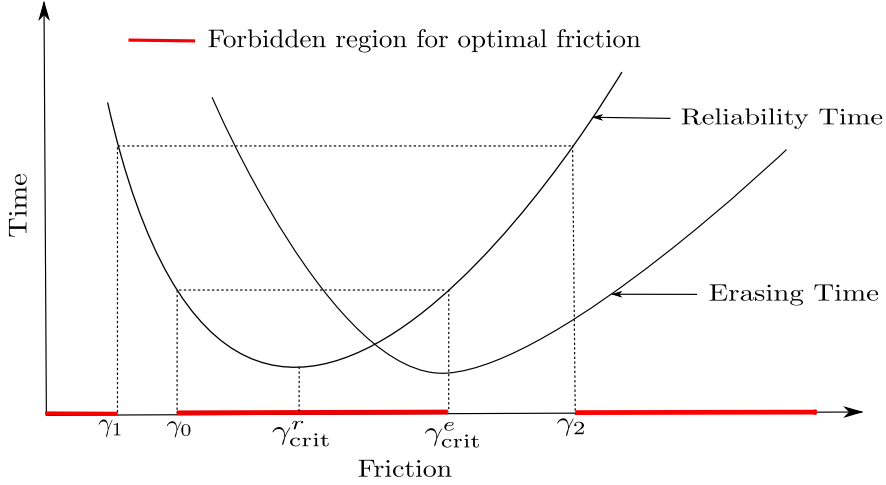

FIGURE 13: Regions of friction-space can be eliminated from the search for optimal bits for our class of controls. As a result, the optimal friction is either critical damping, or lies somewhere within two regions of moderate friction. Illustrative curves of τ_e and τ_r at fixed A, F indicate these regions.

Claim 4.3. [Forbidden regions for optimal friction] Assume that both τ_e and τ_r have a single-well-defined minimum and tend to infinity as γ tends to zero or infinity. Let (A, F, γ) be a (t_r, t_e)-optimal design. (Refer to Figure 13 for notational convenience)

1. Let γ_0 be such that $\tau_r(A, F, \gamma_0) = \tau_r(A, F, \gamma^e_{\text{crit}})$.

   (a) If $\gamma^e_{\text{crit}} > \gamma^r_{\text{crit}}$, then $\gamma \notin (\gamma_0, \gamma^e_{\text{crit}})$.
   (b) If $\gamma^e_{\text{crit}} < \gamma^r_{\text{crit}}$, then $\gamma \notin (\gamma^e_{\text{crit}}, \gamma_0)$.

   i.e. the friction of the optimal bit does not reside in the central red region in Figure 13.

2. Let $A_{\min}$ be the minimum height for a bit to be meaningfully bistable and let $\gamma_1 < \gamma_2$ be such that $\tau_r(A_{\min}, F, \gamma_1) = \tau_r(A_{\min}, F, \gamma_2) = t_r$. If (A, F, γ) is not locally trapped, then $\gamma \notin (0, \gamma_1) \cup (\gamma_2, \infty)$.

   i.e. the friction of the optimal bit does not arise from the extreme red regions in Figure 13.

Proof. 1. We prove it for the case when $\gamma^e_{\text{crit}} > \gamma^r_{\text{crit}}$, the other case proceeds in identical fashion. For contradiction, assume that $\gamma \in (\gamma_0, \gamma^e_{\text{crit}})$. Then, due to the single minima in both τ_e and τ_r, and the fact that τ_r tends to infinity as γ tends to zero or infinity, there exists a design (A, F, γ') with $\gamma' > \gamma_0$ and $\tau_r(A, F, \gamma') = \tau_r(A, F, \gamma) \geq t_r$, but

20

$\tau_e(A, F, \gamma') < \tau_e(A, F, \gamma) \leq t_e$. The design (A, F, γ') is (t_r, t_e)-optimal since it is (t_r, t_e)-feasible and has $W(A, F, \gamma') = W(A, F, \gamma)$, contradicting Lemma 4.1. 1 that the optimal bit saturates the bound on the erasing time constraint.

2. For contradiction, suppose that $\gamma < \gamma_1$ or $\gamma > \gamma_2$. Then since $A \geq A_{\min}$ and the reliability time increases with well height and more extreme values of γ, either $\tau_r(A, F, \gamma) \geq \tau_r(A_{\min}, F, \gamma) > \tau_r(A_{\min}, F, \gamma_1) = t_r$ or $\tau_r(A, F, \gamma) \geq \tau_r(A_{\min}, F, \gamma) > \tau_r(A_{\min}, F, \gamma_2) = t_r$, contradicting claim 4.1. 2 that an optimal design that is not locally trapped saturates the bound on the reliability time constraint.

$\square$

For clarity, let us assume initially that $\gamma_{\text{crit}}^e > \gamma_{\text{crit}}^r$ (equivalent arguments hold for the alternative). We see that optimal designs reside either at γ_{crit}^e, or lie within two regions at moderate friction, as illustrated in Figure 13. Interestingly, one region is adjacent to γ_{crit}^e, whereas the other is not. It is not easy to see how designs in one region ($\gamma_1 \leq \gamma \leq \gamma_0$) as in Figure 13 can outperform those in the other region ($\gamma_{\text{crit}}^e < \gamma \leq \gamma_2$). Indeed, when we performed numerical optimisation on the regression-based fits to our simulation data, we only observed optimal bits that are either critically damped or lie in the allowed region adjacent to critical damping. This is illustrated in Figure 14, where we plot the optimal friction as a function of erasing time requirement at fixed reliability time requirement, for two values of reliability time requirements. We also plot γ_{crit}^e and γ_{crit}^r for comparison. At low erasing time requirements, designs reside at γ_{crit}^e. At slightly higher erasing time requirements, the designs become saturated and the optimal friction lies adjacent to γ_{crit}^e in the region $\gamma_{\text{crit}}^e < \gamma \leq \gamma_2$. Eventually, γ_{crit}^e crosses γ_{crit}^r. At the crossing point, we have $\gamma = \gamma_{\text{crit}}^e = \gamma_{\text{crit}}^r$. At higher values of erasing time requirements, γ still occupies the region adjacent to γ_{crit}^e, which is now $\gamma_1 \leq \gamma \leq \gamma_{\text{crit}}^e < \gamma_{\text{crit}}^r$.

(a) Reliability time requirement$(t_r) = 500$ (b) Reliability time requirement$(t_r) = 10000$
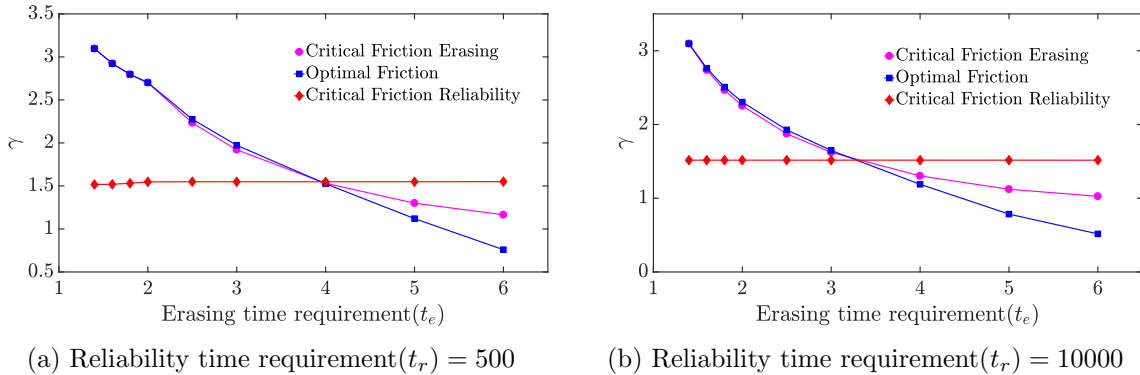
FIGURE 14: Optimal friction is either critical damping, or lies within a small region adjacent to critical damping, for our family of controls. We plot friction for optimal designs (A, F, γ) against erasing time requirements (t_e) for a fixed value of reliability time requirement (t_r), alongside γ_{crit}^e and γ_{crit}^r. Note that A and F are not fixed, but determined by the optimisation procedure alongside the optimal friction for each requirement (t_r, t_e). The data was obtained from numerical optimisation and minimisation based on regression fits to simulation data.

21

5   Conclusions

We have explored the question of the design of optimal bits. Previously, authors have focused on designing optimal protocols that minimize work input when implementing a finite-time operation on a given system [45, 46, 47, 8, 48]. Our approach differs in considering that bits need to have two distinct functionalities: retain data for long periods of time and allow rapid switching or erasing. Moreover we consider optimising over system parameters such as the intrinsic friction as well as the external control. Our fundamental observation is that friction plays a non-trivial role in the design of bits. Both switching/erasing and the eventual degradation of data involve relaxation towards equilibrium from a non-equilibrium distribution. This process is fastest at intermediate values of the friction, but slow in the overdamped and underdamped regimes. The best bit designs have high reliability times and low switching/erasing times, which implies an inherent trade-off in bit design between extreme values of friction that favour high reliability, and moderate values of friction that favour rapid switching or erasing.

We have explored the consequences of the biphasic role of friction for a simple class of controls. The existence of non-trivial minima of erasing time in the level set of work leads to the generation of trapped designs. These designs are optimal for reliability requirements smaller than their own reliability time leading to unsaturated requirements. The result of the trade-off between extreme values of friction that maximize reliability time and moderate values of friction that minimise erasing times is that optimal designs are either critically damped or occupy a region of moderate friction close to critical damping.

Our work opens up a new perspective on the design of efficient computational devices showing that: *the best designs are likely to be neither underdamped nor overdamped.* This observation is particularly important as some authors have considered friction to be inherently problematic for computation [20, 21, 22, 23]. Equally, the role of friction is suppressed when bits are modelled as discrete two-state systems [2, 9, 49], since this approximation assumes rapid equilibration within the discrete states.

We have only considered a simple family of controls to motivate our analysis and illustrate our findings. This family is not optimal - it was chosen for it's simplicity and ease of analysis. Moreover, there is some arbitrariness in the definition of both the erasing and reliability times. As such, the numerical details of the results obtained are not very important. We are not claiming to have derived numerical corrections to the minimal cost of erasing a bit, for example, or the specific work costs (substantially larger than $k_B T \ln 2$) which are not that informative. Rather, it is the qualitative results, which hold for a much broader class of controls which are important. The non-monotonic role of friction in both the erasing and reliability time-scales is a generic physical phenomenon that extends beyond the details of our implementation, and implies a competition between the goals of fast manipulation and long reliability times. Relatively weak assumptions – that it is always possible to decrease work at fixed reliability time and that the minimal erasing time decreases with increased work imply that erasing time requirements are always saturated by optimal bits and that trapped designs lead to unsaturated reliability time requirements respectively. Other results rely more on the simplicity of the control family: the existence of only one local minimum of erasing time at fixed work simplifies the question of whether a requirement specification is saturated. The fact that work is independent of friction simplifies the task of eliminating certain values of friction as sub-optimal.

Explicit exploration of a broader class of controls, including those with more complex

variation over time, and varying parameters such as particle mass and distance between wells, are possible directions for future work. It is not immediately clear whether minima in erasing time at fixed work cost will become more or less prominent features of the optimisation landscape when the complexity of the system is increased, for example. In particular, raising or lowering the barrier between metastable states is a common idea [9, 8, 28, 29]. Lowering the barrier during erasing potentially allows for faster erasing at fixed reliability time and lower work cost. If said barriers could be raised and lowered arbitrarily far and quickly, it may be possible to circumvent any conflict between high reliability and low erasure time. However, real physical systems are not generally this flexible. Indeed, in order to apply a complex time-dependent control to a small colloid, experimenters typically use optical feedback traps [28, 29], which are not true potentials and rely on the continuous input of energy to apply forces and perform feedback control. For true physical protocols that permit finite raising and lowering of barriers between metastable states, we expect that our findings would still apply to a family of protocols with optimal barrier manipulation. An alternative direction would be to consider similar effects in systems with inherently quantum mechanical behaviour.

6 Data accessibility

The codes and computational data for this manuscript can be downloaded from `https://www.imperial.ac.uk/principles-of-biomolecular-systems/contact--obtain-code-and-data/`

7 Competing interests

The authors declare no competing interests.

8 Authors' contributions

MG conceived the project. AD performed the calculations. AD, MG, TEO and NSJ planned the research, analysed the results and wrote the paper.

9 Acknowledgement

10 Funding

References

[1] L. Szilard. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Behavioral Science*, 9(4):301–310, 1964.

[2] R. Landauer. Irreversibility and Heat Generation in the Computing Process. *IBM Journal of Research and Development*, 5(3):183–191, July 1961.

[3] C. H. Bennett. The thermodynamics of computation—a review. *International Journal of Theoretical Physics*, 21(12):905–940, 1982.

[4] C. H. Bennett. Notes on the History of Reversible Computation. *IBM Journal of Research and Development*, 32(1):16–23, January 1988.

[5] C. H. Bennett. Notes on Landauer's principle, reversible computation, and Maxwell's Demon. *Studies in the History and Philosophy of Modern Physics*, 34:501–510, 2003.

[6] M. P. Frank. The Physical Limits of Computing. *Computing in Science and Engineering*, 4(3):16–26, May 2002.

[7] E. Pop. Energy dissipation and transport in nanoscale devices. *Nano Research*, 3(3):147–169, 2010.

[8] P. R. Zulkowski and M. R. DeWeese. Optimal finite-time erasure of a classical bit. *Physical Review E*, 89:052140, May 2014.

[9] T. E. Ouldridge, C. C. Govern, and P. R. Wolde. The thermodynamics of computational copying in biochemical systems. *Phys. Rev. X*, 7:021004, 2017.

[10] J. Von Neumann and A. Burks. Theory of self-reproducing automata. *IEEE Transactions on Neural Networks*, 5(1):3–14, 1966.

[11] R. Sarpeshkar. *Ultra Low Power Bioelectronics: Fundamentals, Biomedical Applications, and Bio-Inspired Systems*. Cambridge, 002 2010.

[12] R. Sarpeshkar. Analog versus Digital: Extrapolating from Electronics to Neurobiology. *Neural computation*, 10(7):1601–1638, 1998.

[13] B. I. Rapoport, J. T. Kedzierski, and R. Sarpeshkar. A glucose fuel cell for implantable brain–machine interfaces. *PloS one*, 7(6):e38436, 2012.

[14] Y. Tu. The nonequilibrium mechanism for ultrasensitivity in a biological switch: Sensing by Maxwell's demons. *Proceedings of the National Academy of Sciences*, 105(33):11737–11741, 2008.

[15] G. Lan, P. Sartori, S. Neumann, V. Sourjik, and Y. Tu. The energy-speed-accuracy trade-off in sensory adaptation. *Nature Physics*, 8(5):422–428, May 2012.

[16] C. C. Govern and P. R. ten Wolde. Optimal resource allocation in cellular sensing systems. *Proceedings of the National Academy of Sciences*, 111(49):17486–17491, 2014.

[17] T. E. Ouldridge and P. R. ten Wolde. Fundamental costs in the production and destruction of persistent polymer copies. *Phys. Rev. Lett.*, 118:158103, 2017.

[18] S. Bo, M. D. Giudice, and A. Celani. Thermodynamic limits to information harvesting by sensory systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(1):P01014, 2015.

[19] A. C. Barato, D. Hartich, and U. Seifert. Efficiency of cellular information processing. *New Journal of Physics*, 16(10):103024, 2014.

[20] W. Anacker. Josephson Computer Technology: An IBM Research Project. *IBM Journal of Research and Development*, 24(2):107–112, March 1980.

[21] M. Büttiker, E. P. Harris, and R. Landauer. Thermal activation in extremely under-damped Josephson-junction circuits. *Physical Review B*, 28:1268–1275, Aug 1983.

[22] M. Klein and A. Mukherjee. Thermal noise induced switching of Josephson logic devices. *Applied Physics Letters*, 40(8):744–747, 1982.

[23] K. K. Likharev. Classical and quantum limitations on energy consumption in computation. *International Journal of Theoretical Physics*, 21(3):311–326, 1982.

[24] P. E. Kloeden and E. Platen. Higher-order implicit strong numerical schemes for stochastic differential equations. *Journal of Statistical Physics*, 66(1):283–314, 1992.

[25] G. A. Pavliotis. *Stochastic processes and applications*, volume 60. Springer, 2014.

[26] J. C. Mattingly and A. M. Stuart. Geometric ergodicity of some hypo-elliptic diffusions for particle motions. *Markov Processes and Related Fields*, 8(2):199–214, 2002.

[27] M. Gopalkrishnan. The Hot Bit I: The Szilard-Landauer Correspondence. *arXiv:1311.3533*, [cs.IT], 2013.

[28] 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(7388):187–189, 2012.

[29] Y. Jun, M. Gavrilov, and J. Bechhoefer. High-Precision Test of Landauer's Principle in a Feedback Trap. *Physical Review Letters*, 113:190601, Nov 2014.

[30] T. Sagawa and M. Ueda. Nonequilibrium thermodynamics of feedback control. *Physical Review E*, 85:021104, Feb 2012.

[31] T. Sagawa and M. Ueda. Generalized Jarzynski Equality under Nonequilibrium Feedback Control. *Physical Review Letters*, 104:090602, Mar 2010.

[32] M. Ponmurugan. Generalized detailed fluctuation theorem under nonequilibrium feedback control. *Physical Review E*, 82:031129, Sep 2010.

[33] D. Abreu and U. Seifert. Thermodynamics of Genuine Nonequilibrium States under Feedback Control. *Physical Review Letters*, 108:030601, Jan 2012.

[34] J. M. Horowitz and S. Vaikuntanathan. Nonequilibrium detailed fluctuation theorem for repeated discrete feedback. *Physical Review E*, 82:061120, Dec 2010.

[35] L. Sourabh, R. Shubhashis, and A.M. Jayannavar. Fluctuation theorems in the presence of information gain and feedback. *Journal of Physics A: Mathematical and Theoretical*, 45(6):065002, 2012.

[36] J. L. Vega, R. Guantes, and S. Miret-Artes. Mean first passage time and the Kramers turnover theory in activated atom-surface diffusion. *Physical Chemistry Chemical Physics*, 4:4985–4991, 2002.

[37] L. D. Ruslan, H. Richard, and M.V. Tretyakov. Langevin thermostat for rigid body dynamics. *The Journal of Chemical Physics*, 130(23):234101, 2009.

[38] K. Sekimoto. Kinetic Characterization of Heat Bath and the Energetics of Thermal Ratchet Models. *Journal of the Physical Society of Japan*, 66(5):1234–1237, 1997.

[39] K. Sekimoto. Langevin Equation and Thermodynamics. *Progress of Theoretical Physics Supplement*, 130:17–27, 1998.

[40] H.A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284 – 304, 1940.

[41] V. I. Melnikov and S. V. Meshkov. Theory of activated rate processes: Exact solution of the Kramers problem. *The Journal of Chemical Physics*, 85(2):1018–1027, 1986.

[42] E. Pollak, H. Grabert, and P. Hänggi. Theory of activated rate processes for arbitrary frequency dependent friction: Solution of the turnover problem. *The Journal of Chemical Physics*, 91(7):4073–4087, 1989.

[43] P. Hänggi, P. Talkner, and M. Borkovec. Reaction-rate theory: Fifty years after Kramers. *Reviews of Modern Physics*, 62:251–341, Apr 1990.

[44] S. Arrhenius. *Über die Dissociationswärme und den Einfluss der Temperatur auf den Dissociationsgrad der Elektrolyte*. Wilhelm Engelmann, 1889.

[45] T. Schmiedl and U. Seifert. Optimal Finite-Time Processes In Stochastic Thermodynamics. *Physical Review Letters*, 98:108301, Mar 2007.

[46] H. Then and A. Engel. Computing the optimal protocol for finite-time processes in stochastic thermodynamics. *Physical Review E*, 77:041105, Apr 2008.

[47] E. Aurell, C. Mejía-Monasterio, and P. Muratore-Ginanneschi. Optimal Protocols and Optimal Transport in Stochastic Thermodynamics. *Physical Review Letters*, 106:250601, Jun 2011.

[48] T. R. Gingrich, G. M. Rotskoff, G. E. Crooks, and P. L. Geissler. Near-optimal protocols in complex nonequilibrium transformations. *Proceedings of the National Academy of Sciences*, 113(37):10263–10268, 2016.

[49] M. Gopalkrishnan. A Cost/Speed/Reliability Tradeoff to Erasing. *Entropy*, 18(5), 2016.

[50] S. Wright and J. Nocedal. Numerical optimization. *Springer Science*, 35:67–68, 1999.

[51] T. Rapcsák. Smooth nonlinear optimization in R^n. *Springer*, 19:376, 1997.

A Appendix

A.1 Validating the timestep of the integrator

We validate the accuracy of our Langevin integrator by considering the dependence of thermodynamic expectations on the time step. We calculate the average potential and kinetic energies for a particle in a quadratic potential $W_{A,B} = A\left(\frac{x}{B} - 1\right)^2$, a quadratic proxy for a single well of the quartic resting-state potential. We plot the results in Figure 15 for a few representative values of the friction coefficient $\gamma = [0.1, 1.10, 100]$ and $A = 10$. Each result is based on an average from 10 simulations each of 5×10^8 time steps. As is evident from the figure, a time step of 0.001 gives good convergence to the equipartition limit of $k_B T / 2$.

(a) $\gamma = 0.1$ (b) $\gamma = 1$

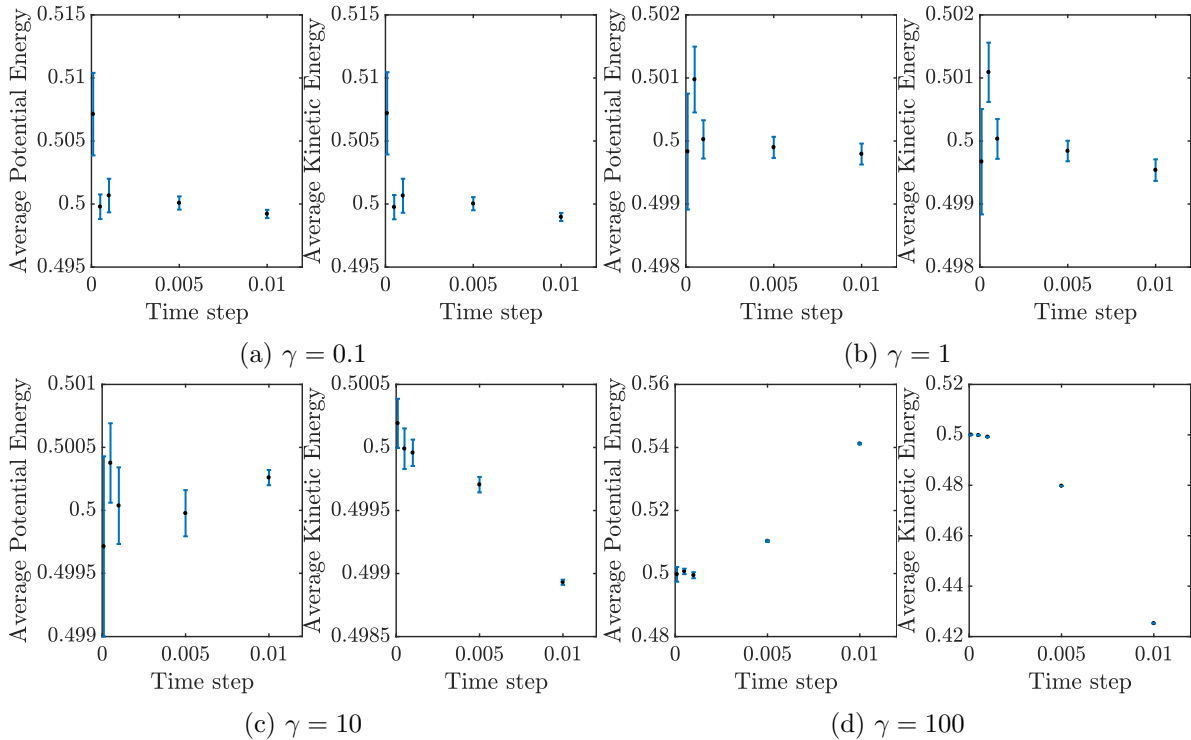(c) $\gamma = 10$ (d) $\gamma = 100$

FIGURE 15: A time step of 0.001 is good enough to ensure that the average potential and kinetic energies approaches $\frac{k_B T}{2} = 0.5$ for a wide range of friction values.

However, it is not sufficient to just compare the average kinetic and potential energies to the equipartition limit. We need to ensure that the observed kinetics are robust to our choice of time step. In particular, we need to test that a time step of 0.001 is sufficient for the highest values of our control parameter F, which presents the most severe challenge to integrating our Langevin equation (due to the behaviour near $x = 0$). Figure 16 confirms that a time step of 0.001 is appropriate for $F = 100$ and the full range of γ tested. Each value in the figure is an average over 1000 initial conditions.

A.2 Erasure region

In this section, we demonstrate that our results are not limited to our specific definition of the erasure region by considering two alternative criteria for erasing and confirming that our
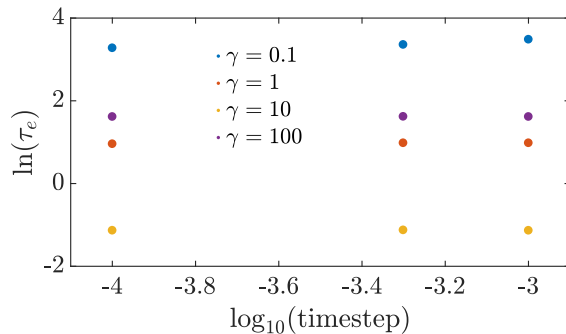
FIGURE 16: A time step of 0.001 gives reasonable values of erasing time for the largest value of the control parameter ($F = 100$) that we use in the simulations.

earlier conclusions are supported.

A.2.1 Accuracy of erasure: Convergence of probability distribution

Recall that the erasing time is a sum of transport time (τ_t) and the mixing time (τ_m). Since the transport time is independent of the metric used to measure the mixing time, we analyse a proxy for the mixing process in isolation. Specifically, we consider the relaxation of a particle in a harmonic well, initially prepared in an arbitrary non-equilibrium distribution. As an alternative definition of mixing, we consider τ_m^ϵ as the first time when the probability distribution of the particle comes within a certain distance (in the appropriate norm, and relative to its initial distribution) of the Gibbs distribution corresponding to the well. More specifically,

$$\tau_m^\epsilon = \inf_{t \geq 0} \{||\text{law}((x(t), p(t)) - \pi_0(x,p)||_{L^2(\pi_0(x,p))} \leq \epsilon ||\text{law}((x(0), p(0)) - \pi_0(x,p)||_{L^2(\pi_0(x,p))} \}$$

where $(x(t), p(t))$ is the solution to Equation 7 given appropriate initial conditions. Here, $\pi_0(x,p)$ is the stationary distribution of the harmonic well. As is usual, the weighted norm $L^2(\pi_0(x,p)) := \{f | \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f|^2 \pi_0(x,p) dx dp < \infty\}$. Informally, τ_e^ϵ is the time required for the distribution to be a factor $\epsilon \ll 1$ "closer" to the equilibrium distribution than in the initial condition.

We define $\eta = 1 - \epsilon$ as the **accuracy** of erasure. Lesser the ϵ, the closer the distribution of the particle is to the Gibbs distribution of the harmonic well and hence more accurate the erasure.

Consider the modified Langevin equation

$$\begin{aligned} m\,dx &= p\,dt \\ dp &= -\gamma p\,dt - \partial_x N_{A,B}(x)\,dt + \sqrt{2m\gamma k_B T}\,dW \end{aligned} \tag{23}$$

28

Here $N_{A,B}(x) = \frac{1}{2} m \omega_0^2 (x - B)^2$ where $\omega_0 = \sqrt{\frac{8A}{mB^2}}$ is the harmonic potential that approximates well "0". Equation23 has the generator [25, pp. 182] given by

$$\mathcal{L} = \frac{p}{m} \partial_x - (\partial_x N_{A,B}(x)) \partial_p + \gamma \left(-p \partial_p + k_B T \partial_p^2 \right) \tag{24}$$

It is common knowledge that the following equation is true [26].

$$||\text{law}((x(t), p(t)) - \pi_0(x,p)||_{L^2(\pi_0(x,p))} \leq e^{-\lambda t} ||\text{law}((x(0), p(0)) - \pi_0(x,p)||_{L^2(\pi_0(x,p))}$$

where λ is the first non-zero eigenvalue of the generator $\mathcal{L}$ given by Equation 24. Setting $e^{-\lambda t} = \epsilon$, we get useful **upper bounds** on the mixing time. In particular, we get

$$\tau_m^\epsilon \leq \frac{1}{\lambda} \ln \frac{1}{\epsilon} \tag{25}$$

For the sake of rough scaling, we will use $\tau_m^\epsilon \approx \frac{1}{\lambda} \ln \frac{1}{\epsilon}$ as an approximate estimate of the mixing time. It is important to note that the generator $\mathcal{L}$ is not self-adjoint and may possess imaginary eigenvalues. The rate of convergence in such cases will be determined by the real part of the eigenvalue. In fact using [25, pp. 200], the first non-zero eigenvalue of the generator is

$$\lambda = \frac{\gamma}{2} - \frac{1}{2} \sqrt{\gamma^2 - 4\omega_0^2}$$

In the underdamped limit when $\gamma \ll 2\omega_0$, we have $Re(\lambda) = \frac{\gamma}{2}$. Therefore $\tau_m^\epsilon \approx \frac{2}{\gamma} \log(\frac{1}{\epsilon})$ in the low friction regime. When friction is very high i.e. $\gamma \gg \omega_0$, we have $\lambda \approx \frac{\omega_0^2}{\gamma}$. As a result we get $\tau_m^\epsilon \approx \frac{\gamma}{\omega_0^2} \log \frac{1}{\epsilon}$. Thus our proxy for the mixing process produces $\tau_m^\epsilon \propto \frac{1}{\gamma}$ in the low friction regime and $\tau_m^\epsilon \propto \gamma$ in the high friction regime, consonant with the scaling and non-monotonicity observed using the erasure region criterion. As a consequence, using a convergence criterion for erasure would not change the physics of the problem, merely perturbing the erasing time-scale quantitatively.

A.2.2 $4k_B T$ criterion for erasure region

Within the framework of the original "erasure region" criterion discussed in the main text, we now consider the robustness of results to changing the numerical value of the criterion. Specifically, we here define the erasure region as all phase space points with total energy atleast $4k_B T$ below the barrier height. More formally,

$$\tau_e^{4k_B T} = \mathbb{E} \left[\inf\{t \geq 0 \mid x(t) < 0 \text{ and } H(x(t), p(t)) \leq A - 4k_B T\} \right]$$

where $(x(t), p(t))$ is the solution to Equation 7 with the initial condition $(x(0), p(0)) \sim_{\text{law}} \pi_1(x,p)$. We now show that we get the same non-monotonicity and scaling of erasing time as a function of friction-coefficient that we got using the $3k_B T$ criterion. In particular, the erasing time scales as $\frac{1}{\gamma}$ in the low friction regime and scales as γ at high friction. Figure 17 illustrates this fact. Fits are performed using analytical expressions equivalent to those discussed in the main text, but adjusted for the new numerical value of the boundary of the erasure region.

1. Low friction regime:

$$\tau_e^{4k_B T} \approx \sqrt{\frac{2mB}{F}} + \frac{1}{\gamma} \ln \frac{A + F \cdot B}{A - 4k_B T}. \tag{26}$$

2. High friction regime:

$$\tau_e^{4k_B T} \approx \frac{mB\gamma}{F} + \frac{2mB^2\gamma}{5A} \tag{27}$$
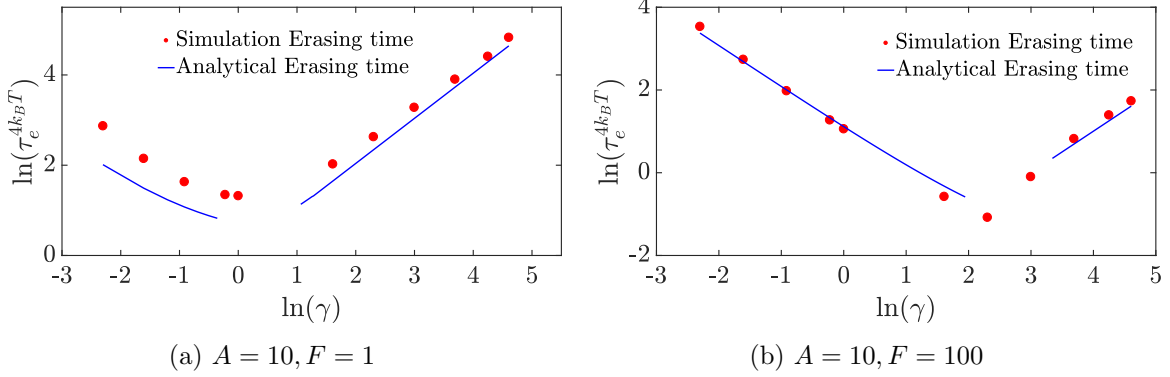

(a) $A = 10, F = 1$ (b) $A = 10, F = 100$

FIGURE 17: Evidence from simulation that the use of $4k_B T$ to define the erasure reason does not change the fundamental physics of the problem.

A.3 Work calculation

A.3.1 Full calculation for work done at t=0

We first provide a more detailed justification of the approximation $W = (A + F \cdot B - k_B T/2)/2$ for the work done when the control potential is switched on, then compare it to the exact result. From Equation 12, and letting $I = \{x \geq 0 \mid A - U_{A,B}(x) + F \cdot x \geq 0\}$, we get

$$
\begin{aligned}
\langle W \rangle &= \int_0^\tau \int_I V_F(x, t)\, p(x, t)\, \delta(t)\, dx dt \\
&= \int_I (A - U_{A,B}(x) + F \cdot x)\, p(x, 0)\, dx
\end{aligned}
\tag{28}
$$

Since $p(x, 0) \propto e^{-\frac{U_{A,B}(x)}{k_B T}}$, we can rewrite the expression of work as

$$
\langle W \rangle = \frac{\int_I (A - U_{A,B}(x) + F \cdot x)\, e^{-\frac{U_{A,B}(x)}{k_B T}}\, dx}{\int_{-\infty}^\infty e^{-\frac{U_{A,B}(x)}{k_B T}}\, dx}
\tag{29}
$$

30

Since $I \subseteq [0, \infty)$ and $\left(A - U_{A,B}(x) + F \cdot x\right) e^{-\frac{U_{A,B}(x)}{k_B T}}$ is negligible as $x \to \infty$, replacing the upper limit of integration by ∞ is reasonable. Hence the integral becomes

$$\langle W \rangle \approx \frac{\int\limits_0^\infty \left(A - U_{A,B}(x) + F \cdot x\right) e^{-\frac{U_{A,B}(x)}{k_B T}} \, dx}{\int\limits_{-\infty}^\infty e^{-\frac{U_{A,B}(x)}{k_B T}} \, dx} \tag{30}$$

When $A >> k_B T$, we can use Bessel's functions to approximate this integral giving

$$\langle W \rangle \approx \frac{\left(A + F \cdot B - \frac{k_B T}{2}\right)}{2} \tag{31}$$

justifying the crude approximation in the main text. The accuracy of this expression compared to Eq. 29 is illustrated in Figure 18.
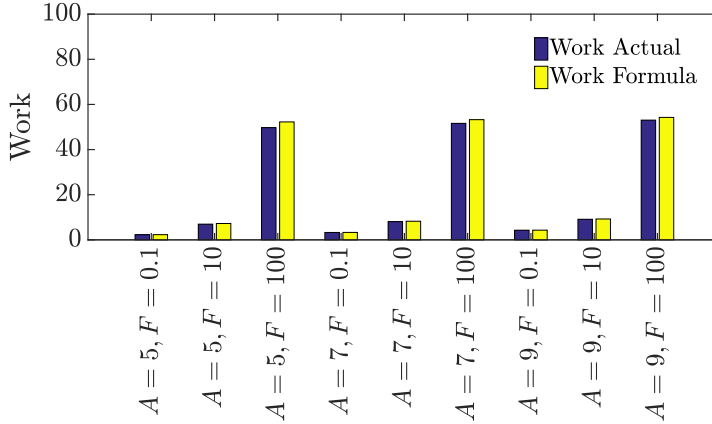

FIGURE 18: Work when the control is switched on against various values of A and F, comparing the full expression Eq. 29 and the approximate result Eq. 31.

A.4 The potential for energy recovery is negligible

Here, we argue that energy recoverable at the end of the protocol is very small, and hence may be neglected. We assume that the control is switched off after a time τ sufficiently large compared to τ_e so that the proportion of particles remaining on the right hand side of the well is determined by the Boltzmann factor. The work that we could then in principle recover is given by the following expression:

$$\langle W_{rec} \rangle \approx \frac{\int\limits_I \left(A - U_{A,B}(x) + F \cdot x\right) e^{\frac{-(A + F \cdot x)}{k_B T}} \, dx}{\int\limits_{-\infty}^0 e^{\frac{-U_{A,B}(x)}{k_B T}} \, dx + \int\limits_0^\infty e^{\frac{-(A + F \cdot x)}{k_B T}} \, dx} \tag{32}$$

31

Recall that $I = \{x \geq 0 \mid A - U_{A,B}(x) + F \cdot x \geq 0\}$. This implies that $I = [0, x^*]$, where $A - U_{A,B}(x^*) + F \cdot x^* = 0$. Using Equations 31 and 32, we will calculate the fraction of recovered work i.e., $W_{rec}^f = \frac{\langle W_{rec} \rangle}{\langle W \rangle}$. Figure 19 precisely calculates this quantity. As is evident from the figure, the fraction is almost negligible and reaches its maximum value at low A and F.
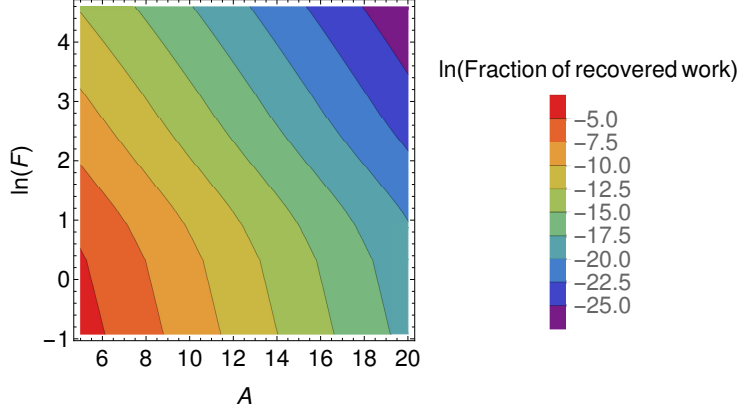

FIGURE 19: Negligible energy can be recovered for our family of controls.

A.5 Regression and Cross Validation

We use cubic regression to interpolate between simulation data for both the reliability and erasing time-scales. Let $F' = \log(F)$ and $\gamma' = \log(\gamma)$. Then we use the following polynomials to fit the time-scales.

1. **Erasing Polynomial**:

$$
\begin{aligned}
\log(\tau_e) = {}& b_1 + b_2 A^3 + b_3 F'^3 + b_4 \gamma'^3 + b_5 A^2 F' + b_6 A' F'^2 + b_7 F'^2 \gamma' + b_8 F' \gamma'^2 \\
& + b_9 A'^2 \gamma' + b_{10} A \gamma'^2 + b_{11} A^2 + b_{12} F'^2 + b_{13} \gamma'^2 + b_{14} AF' + b_{15} F' \gamma' \\
& + b_{16} A \gamma' + b_{17} A + b_{18} F' + b_{19} \gamma'
\end{aligned} \tag{33}
$$

2. **Reliability Polynomial**:

$$
\log(\tau_r) = c_1 + c_2 A^3 + c_3 \gamma'^3 + c_4 A^2 \gamma' + c_5 A \gamma'^2 + c_6 A^2 + c_7 \gamma'^2 + c_8 A \gamma' + c_9 A + c_{10} \gamma \tag{34}
$$

, where the coefficients $b_1, b_2, \cdots b_{19}$ and $c_1, c_2, \cdots c_{10}$ are to be determined by regression.

Figure 20 gives a visual illustration of the fact that cubic fits offer a good approximation to the simulation results for both the erasing and reliability time-scales. In what follows, we present a more detailed and formal justification using cross-validation.

We perform "Leave-one-out" cross validation to justify the use of cubic regression. Figure 21 reports the mean square training and testing cross-validation errors corresponding to linear, quadratic and cubic fits. A lower value of the testing error indicates a good fit. Cubic regression has the lowest value of testing errors amongst the fits considered for both
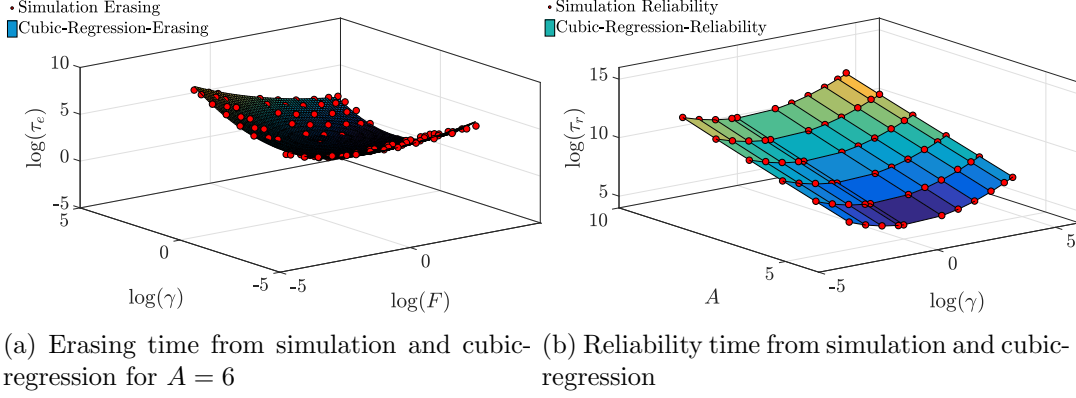
(a) Erasing time from simulation and cubic-regression for $A = 6$

(b) Reliability time from simulation and cubic-regression

FIGURE 20

(a) Errors of regression fits for erasing time

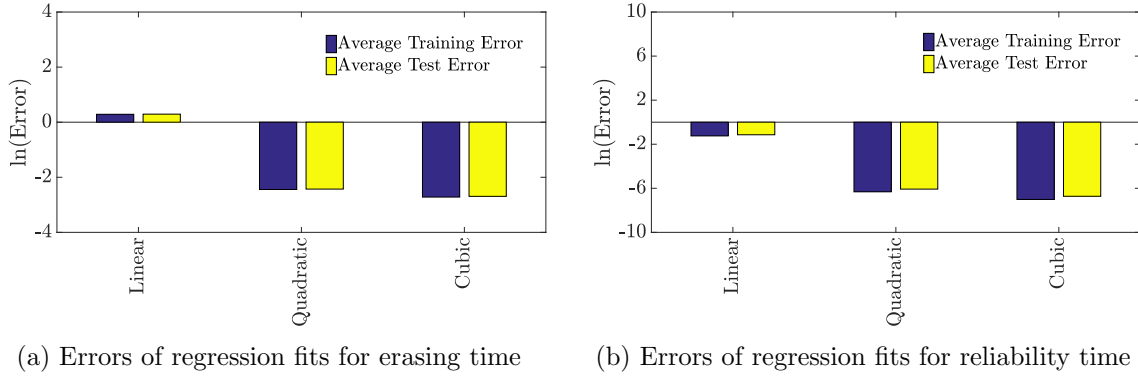(b) Errors of regression fits for reliability time

FIGURE 21

the reliability and erasing time-scales. Figure 21 confirms that the training and testing errors corresponding to cubic-regression for both the time-scales are roughly comparable(with the training error being slightly lower than the testing error). As a result, we can safely assume that the cubic polynomial does not over-fit the data and use it for modelling both the time-scales.

A.6 Calculation of well parameters

Here we calculate the quantities needed to apply Equation 13 to our system, with the potential $U_{A,B}(x) = A\left(\frac{x^2}{B^2} - 1\right)^2$. We have $\partial_x U_{A,B}(x) = \frac{4Ax(x^2 - B^2)}{B^4}$ and $\partial_{xx} U_{A,B}(x) = \frac{4A(3x^2 - B^2)}{B^4}$.

1. **Angular frequency at barrier height** (ω_b): We can approximate the region near the barrier by an inverted harmonic oscillator. By Taylor expanding the potential about the point $x = 0$, we get

$$U_{A,B}(x) \approx U(0) + \partial_x U_{A,B}(x)\Big|_{x=0} x + \frac{\partial_{xx} U_{A,B}(x)\Big|_{x=0} x^2}{2} \approx A - \frac{2Ax^2}{B^2} = A - \frac{m\omega_b^2 x^2}{2} \tag{35}$$

Therefore we have $\omega_b = \sqrt{\frac{4A}{mB^2}}$.

33

2. **Angular frequency at the bottom of the well** (ω_0): We can approximate the region near the bottom of the well by a harmonic oscillator. By Taylor expanding the potential about the point $x = B$, we get

$$U_{A,B}(x) \approx U(B) + \partial_x U_{A,B}(x)\Big|_{x=B}(x-B) + \frac{\partial_{xx} U_{A,B}(x)\Big|_{x=B}(x-B)^2}{2} \qquad (36)$$
$$= \frac{4A(x-B)^2}{B^2} = \frac{m\omega_0^2(x-B)^2}{2}$$

Therefore we have $\omega_0 = \sqrt{\frac{8A}{mB^2}}$.

3. **Action at barrier height** $I(A)$: Consider a particle of mass m with a starting velocity $v = 0$, moving along a constant energy surface with energy A. The particle starts at $x = 0$ and moves to $x = \sqrt{2}B$ and returns back to $x = 0$. The action for this round trip is given by $I(A) = \oint p dx = 2\sqrt{2m}\int_0^{\sqrt{2}B}\sqrt{A - A\left(\frac{x^2}{B^2} - 1\right)^2}dx = \frac{8B\sqrt{mA}}{3}$.

A.7 Locally trapped bits are uniquely trapped

In this section, we give typical plots for our family of controls that show no evidence of multiple local minima in erasing time within a level set of work. Towards this we let $F' = \log(F)$ and $\gamma' = \log(\gamma)$. Using the same form of regression polynomial as in Equation 33, but at constant work W, this translates to

$$\log(\tau_e) = b_1 + b_2(W - e^{F'})^3 + b_3 F'^3 + b_4\gamma'^3 + b_5(W - e^{F'})^2 F' + b_6(W - e^{F'})F'^2 + b_7 F'^2\gamma' + b_8 F'\gamma'^2$$
$$+ b_9(W - e^{F'})^2\gamma' + b_{10}(W - e^{F'})\gamma'^2 + b_{11}(W - e^{F'})^2 + b_{12}F'^2 + b_{13}\gamma'^2 + b_{14}(W - e^{F'})F' \qquad (37)$$
$$+ b_{15}F'\gamma' + b_{16}(W - e^{F'})\gamma' + b_{17}(W - e^{F'}) + b_{18}F' + b_{19}\gamma'$$

Note that $\left(\frac{d\tau_e}{d\gamma'}\right)_{W,F'} = \gamma\left(\frac{d\tau_e}{d\gamma}\right)_{W,F}$ and $\left(\frac{d\tau_e}{dF'}\right)_{W,\gamma'} = F\left(\frac{d\tau_e}{dF}\right)_{W,\gamma}$. Therefore solving for $\left(\frac{d\tau_e}{d\gamma}\right)_{W,F} = \left(\frac{d\tau_e}{dF}\right)_{W,\gamma} = 0$ is equivalent to solving for $\left(\frac{d\tau_e}{d\gamma'}\right)_{W,F'} = \left(\frac{d\tau_e}{dF'}\right)_{W,\gamma'} = 0$. We solve $\left(\frac{d\tau_e}{d\gamma'}\right)_{W,F'} = \left(\frac{d\tau_e}{dF'}\right)_{W,\gamma'} = 0$ numerically and plot it in Figure 22. As illustrated by Figure 22, there is exactly one solution to the equations $\left(\frac{d\tau_e}{d\gamma}\right)_{W,F} = \left(\frac{d\tau_e}{dF}\right)_{W,\gamma} = 0$ within the broad range of parameters allowed, confirming our assumption that locally trapped bits are uniquely trapped.

A.8 Geometry of the Optimal Bit

1. **Proof of Claim 4.1**

   Proof. (a) For contradiction suppose that $\tau_e(A, F, \gamma) < t_e$. Since τ_e is continuous and it is possible to locally decrease work at fixed reliability time, there exists a design (A', F', γ') with $W(A', F', \gamma') < W(A, F, \gamma)$ that is (t_r, t_e)-feasible contradicting the optimality of the design (A, F, γ).
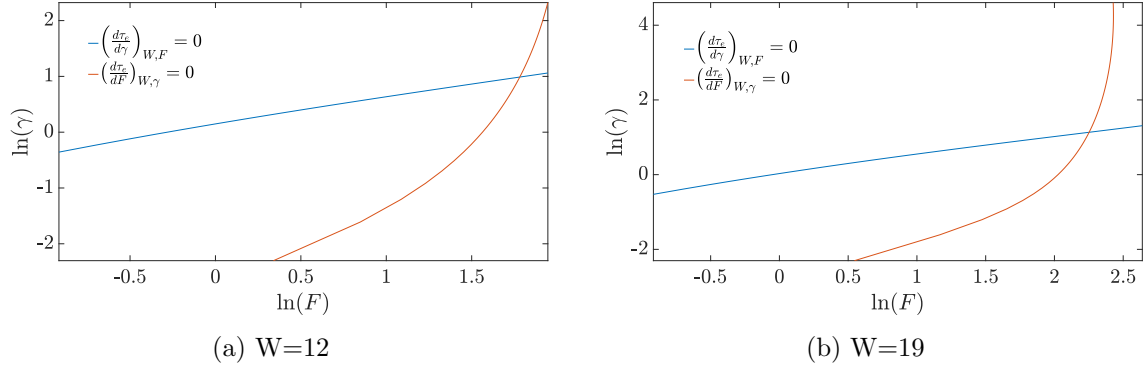
34

(a) W=12   (b) W=19

FIGURE 22: Evidence that for our family of controls, locally trapped bits are uniquely trapped. The system $\left(\frac{d\tau_e}{d\gamma}\right)_{W,F} = \left(\frac{d\tau_e}{dF}\right)_{W,\gamma} = 0$ has exactly one solution within the broad range of parameters considered, which corresponds to a unique local minimum of erasing time in a level set of work. This is illustrated for work $W = 12$ and $W = 19$. The situation is representative for other values of work.

(b) For contradiction suppose that $\tau_r(A, F, \gamma) > t_r$. Since the design (A, F, γ) is not locally trapped and τ_r is continuous, there exists a design (A_0, F_0, γ_0) requiring work $W(A_0, F_0, \gamma_0) = W(A, F, \gamma)$, erasing time $\tau_e(A_0, F_0, \gamma_0) < \tau_e(A, F, \gamma) \leq t_e$ and maintaining reliability time $\tau_r(A_0, F_0, \gamma_0) \geq t_r$. Thus the design (A_0, F_0, γ_0) is (t_r, t_e)-optimal contradicting Claim 4.1.1 that the optimal design saturates the bound on the erasing time constraint.

$\square$

2.

Observation A.1. The erasing time of trapped designs is a strictly decreasing function of work for our family of protocols (Example 2.1). In other words, if (A_0, F_0, γ_0) and (A^*, F^*, γ^*) are trapped designs with $W(A^*, F^*, \gamma^*) > W(A_0, F_0, \gamma_0)$ then $\tau_e(A^*, F^*, \gamma^*) < \tau_e(A_0, F_0, \gamma_0)$.

Proof. Since W is a continuous increasing function of A (Observation 2.2), one can choose $A' > A^*$ such that $W(A', F_0, \gamma_0) = W(A^*, F^*, \gamma^*)$. Noting that increasing well height at fixed F and γ decreases erasing time (Observation 3.1), we get $\tau_e(A', F_0, \gamma_0) < \tau_e(A_0, F_0, \gamma_0)$. Using the fact that (A^*, F^*, γ^*) is a trapped design, we get $\tau_e(A^*, F^*, \gamma^*) \leq \tau_e(A', F_0, \gamma_0) < \tau_e(A_0, F_0, \gamma_0)$ establishing the claim. $\square$

3. **Proof of Claim 4.2**

Proof. (a) Since $\tau_r(A^*, F^*, \gamma^*) \geq t_r$ and $\tau_e(A^*, F^*, \gamma^*) = t_e$, the design (A^*, F^*, γ^*) is (t_r, t_e)-feasible. Suppose that the design (A^*, F^*, γ^*) is not (t_r, t_e)-optimal. Then there exists a (t_r, t_e)-feasible design (A', F', γ') such that $W(A', F', \gamma') < W(A^*, F^*, \gamma^*)$. Let (A_0, F_0, γ_0) be a trapped design with $W(A_0, F_0, \gamma_0) = W(A', F', \gamma') < W(A^*, F^*, \gamma^*)$. Then $\tau_e(A_0, F_0, \gamma_0) > \tau_e(A^*, F^*, \gamma^*)$ since the erasing time of

trapped bits is a strictly decreasing function of work. Using the fact that (A_0, F_0, γ_0) is a trapped design, we get $\tau_e(A', F', \gamma') \geq \tau_e(A_0, F_0, \gamma_0) > \tau_e(A^*, F^*, \gamma^*) = t_e$, a contradiction since (A', F', γ') is a (t_r, t_e)-feasible design.

(b) Immediate from Claim 4.2. 1.

(c) For contradiction suppose that the requirement (t_r, t_e) is unsaturated. Then by Claim 4.1. 1 there exists a (t_r, t_e)-optimal design (A_0, F_0, γ_0) such that $\tau_r(A_0, F_0, \gamma_0) > t_r$ and $\tau_e(A_0, F_0, \gamma_0) = \tau_e(A^*, F^*, \gamma^*) = t_e$. Since locally trapped designs are uniquely trapped, using Claim 4.1. 2, we get that the design (A_0, F_0, γ_0) must be uniquely trapped. Noting that uniquely trapped bits are trapped and using the fact that the erasing time of trapped designs is a strictly decreasing function of work, we get $W(A_0, F_0, \gamma_0) = W(A^*, F^*, \gamma^*)$. This implies that $(A_0, F_0, \gamma_0) = (A^*, F^*, \gamma^*)$, a contradiction since $\tau_r(A_0, F_0, \gamma_0) > t_r \geq \tau_r(A^*, F^*, \gamma^*)$.

$\square$

A.9   Alternative proof via KKT conditions

KKT conditions form the foundation of optimization problems [50, 51]. In order to study the KKT conditions, we consider the optimization problem of finding the design with the lowest work that is (t_r, t_e)-feasible.

Problem A.2.

$$(A^*, F^*, \gamma^*) = \arg \inf_{A, F, \gamma} W(A, F)$$

$$t_r - \tau_r(A^*, \gamma^*) \leq 0$$

$$\tau_e(A^*, F^*, \gamma^*) - t_e \leq 0$$

In order to state the KKT conditions, we will need the notion of a regular point. The following definition will make this precise.

Definition A.3 (**Regular point**). Let $Sat(x^*)$ denote the set of gradients of the constraints that are saturated at the point x^*. Then x^* is regular iff $Sat(x^*)$ does not form a linearly dependent set.

Theorem A.4 (**KKT conditions**). *Let (A^*, F^*, γ^*) be a local optimum of A.2 and a regular point. Then by [50, (12.1), pp. 95], there exists $\lambda_1^*, \lambda_2^* \in \mathbb{R}_{\geq 0}$ such that*

1. $\nabla W(A^*, F^*, \gamma^*) - \lambda_1^* \nabla \tau_r(A^*, \gamma^*) + \lambda_2^* \nabla \tau_e(A^*, F^*, \gamma^*) = 0$.

2. $\lambda_1^* (t_r - \tau_r(A^*, \gamma^*)) = 0$ *and* $\lambda_2^* (\tau_e(A^*, F^*, \gamma^*) - t_e) = 0$.

Given this powerful theorem A.4, we are now ready to prove the the same result that we obtained earlier but with the machinery of KKT conditions.

Lemma A.5. *Let us assume that it is always possible to locally decrease work at fixed reliability time. Let (A^*, F^*, γ^*) be a local optimum of A.2. Then either*

1. *The design (A^*, F^*, γ^*) saturates the bound on both constraints i.e. $\tau_r(A^*, \gamma^*) = t_r$ and $\tau_e(A^*, F^*, \gamma^*) = t_e$ or*

2. The design (A^*, F^*, γ^*) saturates the bound on the erasing time constraint i.e. $\tau_e(A^*, F^*, \gamma^*) = t_e$ but does not saturate the bound on the reliability time constraint i.e. $\tau_r(A^*, \gamma^*) > t_r$ and is locally trapped.

Proof. Consider an optimal design (A^*, F^*, γ^*) such that either it does not saturate the bound on the reliability time constraint i.e. $\tau_r(A^*, \gamma^*) > t_r$ or it does not saturate the bound on the erasing time constraint i.e. $\tau_e(A^*, F^*, \gamma^*) < t_e$. Then we have the following cases:

- Case 1: The design (A^*, F^*, γ^*) saturates the bound on the erasing time constraint, but does not saturate the bound on the reliability time constraint i.e. $\tau_r(A^*, \gamma^*) > t_r$ and $\tau_e(A^*, F^*, \gamma^*) = t_e$. This implies that $\lambda_1^* = 0$. Since only one constraint is active, (A^*, F^*, γ^*) is a regular point. Hence, by Theorem A.4 on KKT conditions, there exists $\lambda_2^* > 0$ such that $\nabla W(A^*, F^*, \gamma^*) + \lambda_2^* \nabla t_e(A^*, F^*, \gamma^*) = 0$. This implies that (A^*, F^*, γ^*) is a stationary point of erasing time in the level set of it's work $W(A^*, F^*, \gamma^*)$. The fact that this stationary point is actually a local minimum follows from Claim 4.1. 2.

- Case 2: The design (A^*, F^*, γ^*) saturates the bound on the reliability time constraint, but does not saturate the bound on the erasing time constraint i.e. $\tau_r(A^*, \gamma^*) = t_r$ and $\tau_e(A^*, F^*, \gamma^*) < t_e$. This implies that $\lambda_2^* = 0$. Since only one constraint is active, (A^*, F^*, γ^*) is a regular point. Hence, by Theorem A.4 on KKT conditions, there exists $\lambda_1^* > 0$ such that $\nabla W(A^*, F^*, \gamma^*) = \lambda_1^* \nabla t_r(A^*, \gamma^*)$, a contradiction since $\frac{\partial W}{\partial F} \neq 0$ but $\frac{\partial \tau_r}{\partial F} = 0$.

- Case 3: The design (A^*, F^*, γ^*) does not saturate the bound on either constraints i.e. $\tau_r(A^*, \gamma^*) > t_r$ and $\tau_e(A^*, F^*, \gamma^*) < t_e$. Since no constraint is active we have $\nabla W(A^*, F^*, \gamma^*) = 0$, which is not possible.

□