

# Direct measurement of nonequilibrium system entropy is consistent with Gibbs-Shannon form

Momčilo Gavrilov<sup>1,†</sup>, Raphaël Chétrite<sup>1,2,3</sup>, and John Bechhoefer<sup>1\*</sup>

<sup>1</sup> Department of Physics, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada

<sup>2</sup> Pacific Institute for the Mathematical Sciences,  
UMI 3069, Vancouver, British Columbia, Canada

<sup>3</sup> Université Côte d'Azur, CNRS, LJAD, Parc Valrose, 06108 NICE Cedex 02, France

†Present address: Department of Biophysics and Biophysical Chemistry,  
Johns Hopkins University, 725 N. Wolfe Street, Baltimore, MD 21205-2185, USA

Stochastic thermodynamics extends classical thermodynamics to small systems in contact with one or more heat baths. It can account for the effects of thermal fluctuations and describe systems far from thermodynamic equilibrium. A basic assumption is that the expression for Shannon entropy is the appropriate description for the entropy of a nonequilibrium system in such a setting. Here, for the first time, we measure experimentally this function. Our system is a micron-scale colloidal particle in water, in a virtual double-well potential created by a feedback trap. We measure the work to erase a fraction of a bit of information and show that it is bounded by the Shannon entropy for a two-state system. Further, by measuring directly the reversibility of slow protocols, we can distinguish unambiguously between protocols that can and cannot reach the expected thermodynamic bounds.

## INTRODUCTION

Beginning with the foundational work of Clausius, Maxwell, and Boltzmann in the 19th c., the concept of entropy has played a key role in thermodynamics. Yet, despite its importance, entropy is an elusive concept [1–8], with no unique definition; rather, the appropriate definition of entropy depends on the scale, relevant thermodynamic variables, and nature of the system, with ongoing debate existing over the proper definition even for equilibrium cases [9]. Moreover, entropy has not been directly measured but is rather inferred from other quantities, such as the integral of the specific heat divided by temperature. Here, by measuring the work required to erase a fraction of a bit of information, we isolate directly the change in entropy in an open nonequilibrium system, showing that it is compatible with the functional form proposed by Gibbs and Shannon, giving it a physical meaning in this context. Knowing the relevant form of entropy is crucial for efforts to extend thermodynamics to systems out of equilibrium.

For a continuous classical system whose state in phase space  $x$  is distributed as the probability density function  $\rho(x)$ , the Gibbs-Shannon entropy is [10, 11]

$$S = -k_B \int dx \rho(x) \ln \rho(x), \quad (1)$$

where  $k_B$  is Boltzmann's constant. For quantum systems, von Neumann introduced, in 1927, the corresponding expression in terms of the density matrix [12]. Historically, the system in Eq. 1 has typically been assumed to be in thermal equilibrium.

The physical relevance of Eq. 1 for a nonequilibrium distribution  $\rho(x)$  has often been questioned (e.g., [13–17]). One concern is that  $S$  is constant on an isolated Hamiltonian system and can change only when evaluated on subsystems, such as those picked out by coarse graining. With many ways to choose subsystems or to coarse grain, is the associated notion of irreversibility intrinsic to the description of the system?

In another approach to entropy, advanced in the context of communication and information theory, Shannon [11, 18] proved that, up to a multiplicative constant,  $S$  is the only possible function satisfying three intuitive axioms. Alternatively, one can start from an axiomatic framework for thermodynamics [19, 20]. The importance of using the appropriate form of entropy is highlighted in the recently developed field of *stochastic thermodynamics* [21–30], where a central, underlying hypothesis is that Eq. 1 applies to densities defined for nonequilibrium mesoscopic systems coupled to one or more heat baths [27]. We emphasize that this extension of the equilibrium Gibbs-Shannon entropy to nonequilibrium systems remains controversial within part of the statistical physics community, mainly for the reason that it is constant for Hamiltonian systems.

In this letter, we offer an experimental approach: in a nonequilibrium system, we measure directly the change in the entropy of the system and show that it is compatible with the postulated Gibbs-Shannon form, Eq. 1.

Our system is a micron-scale silica bead in water at temperature  $T$  that serves as a reservoir, or heat bath. We use a feedback trap [31] to create a virtual symmetric double-well potential  $U(x, t)$  that models a one-bit memory. The particle motion in this trap obeys nearly ideal

\* email: johnb@sfu.ca

Langevin (Brownian) overdamped dynamics [32–34],

$$\dot{x}(t) = -\frac{1}{\gamma} \frac{\partial U(x, t)}{\partial x} \Big|_{x(t)} + \sqrt{\frac{2k_B T}{\gamma}} \nu(t), \quad (2)$$

where  $\nu(t)$  denotes white-noise forcing, Gaussian with unit variance, and  $\gamma$  denotes the damping.

We show that erasing a fraction of this bit requires, from a generalization of the Landauer principle for a two-state system [35], a minimal average work whose value is set by the Gibbs-Shannon system entropy given in Eq. 1. Our main goal, however, is not to further explore Landauer’s principle but rather to *use* it to test whether the Shannon entropy has a physical meaning in the nonequilibrium contexts probed by our experiments.

Appropriate experimental protocols require complex, precise control of the shape of the potential,  $U(x, t)$ . Such control—Involving barrier height, tilt, and local coordinate stretching to produce asymmetry between macrostates—is easy to achieve using feedback traps, where the form of a “virtual potential” is defined in software by applying the force that would be applied by a physical potential (see Methods). By contrast, it is very difficult to achieve using an ordinary, physical potential. Combining those operations, we construct thermodynamically reversible protocols that can reach theoretical bounds for required work in the slow limit.

## THEORY

### Second law of thermodynamics in terms of work

The second law of thermodynamics asserts that during a time interval  $[0, \tau]$ , the entropy production  $S_{\text{tot}} \geq 0$  [2, 25, 36–38]. This entropy production is that of the total system, including the surrounding medium (heat bath), and decomposes into two terms:

$$S_{\text{tot}} = S_m + \Delta S, \quad (3)$$

where  $S_m$  is the entropy exchanged with the surrounding medium and where  $\Delta S = S_\tau - S_0$  is the difference in system entropy over the time interval. At this point,  $S_0$  and  $S_\tau$  are not necessarily given by the Shannon entropy. Using the Clausius principle (1850) for the equilibrium bath [39], we can write  $S_m = Q/T$ , where  $Q$  is the heat exchanged with the medium, defined to be positive if the transfer is from the system to the medium. Mathematically, the equilibrium character of the bath is reflected by the fact that the amplitude  $2k_B T$  in front of the noise term in the Langevin equation, Eq. 2, is constant and well defined during the entire protocol. Physically, this hypothesis means that the time scales of the particle are much slower than those of the bath.

The second law then becomes  $Q \geq -T \Delta S$ . To reformulate the second law in terms of work, we use the first

law,

$$W = \Delta E + Q, \quad (4)$$

where  $W$  is the average work done on the system to carry out the protocol over time  $\tau$ . In the context of stochastic thermodynamics for overdamped dynamics, Eq. 2—small systems in contact with a large bath—work is calculated using the average of the Sekimoto formula (Eq. 16 in methods) [21, 26, 29, 40]. Then, using the nonequilibrium free energy  $F_{\text{neq}} = E - TS$ , the expression for heat  $Q$  given above, and Eq. 4, we have [41, 42]

$$W \geq \Delta F_{\text{neq}}. \quad (5)$$

Note that the average energy  $E$  at time  $t$  is determined from the potential  $U(x, t)$  and the instantaneous density of the process  $\rho(x, t)$  by

$$E(t) = \int_{-\infty}^{\infty} dx \rho(x, t) U(x, t). \quad (6)$$

The nonequilibrium free energy  $F_{\text{neq}}$  reduces to the conventional equilibrium free energy, defined using the partition function, when the average energy  $E$  and entropy  $S$  are evaluated from equilibrium distributions.

### Coarse graining from a continuous to a discrete system

In our experiments, we measure the continuous position  $x(t)$  in a double-well symmetric potential  $U(x, t)$ . Because the energy barrier  $E_b$  of the double-well potential is much higher than  $k_B T$  for initial and final states, we can consider the system to be effectively a two-state system at those times, with the particle either in the left well (state  $L$ ), defined by  $x < 0$ , or the right well (state  $R$ ), defined by  $x > 0$ . In this section, we derive the second law for such initial/final two-state systems, relating it explicitly to the underlying continuum description.

To accomplish this, we define the notion of local equilibrium in the potential  $U(x, t)$ , where, in the discussion below,  $t$  is either the initial time 0 or the final time  $\tau$ . That is the system is in state  $L$  (left) with probability  $p(t)$  and state  $R$  (right) with probability  $1 - p(t)$ . But, constrained to be within one well or the other, the system is in thermal equilibrium.

We can thus define a conditional equilibrium free energy  $F_{\text{leq}}(t)$ , which is the free energy of the system given that it is in the left well [43, 44]. In analogy with the usual definition of the equilibrium free energy, we have,

$$F_{\text{leq}}(t) = -k_B T \ln Z_{\text{leq}}(t), \quad (7)$$

where the conditional partition function  $Z_{\text{leq}}(t)$  is given by integrating  $\exp[-U(x, t)/k_B T]$  over the interval  $(0, \infty)$ .  $F_{\text{leq}}(t)$  is also known as the “conformational”

free energy [45]. Because of the assumed symmetry of the initial/final potential,  $F_{\text{leq}}(t)$  is the same if evaluated over the other state,  $R$ . Otherwise, one would define local quantities for each state. Notice that we can invert Eq. 7 to write  $Z_{\text{leq}}(t) = \exp[-F_{\text{leq}}(t)/(k_B T)]$ .

We can then define a local-equilibrium density function,

$$\rho_{\text{leq}}(x, t) = \exp [(F_{\text{leq}}(t) - U(x, t))/(k_B T)] \times [p(t) \theta(-x) + [1 - p(t)] \theta(x)], \quad (8)$$

where  $\theta(x)$  is the Heaviside step function, 0 for  $x < 0$  and 1 for  $x \geq 0$ . The physical meaning of  $\rho_{\text{leq}}(x, t)$  is that the particle is in local equilibrium in the left well of the potential  $U(x, t)$  with probability  $p(t)$  and in local equilibrium in the right well with probability  $1 - p(t)$ .

Notice, too, that  $\rho_{\text{leq}}(x, t)$  is typically not the global equilibrium Boltzmann-Gibbs distribution associated with the potential  $U(x, t)$ , which would have  $p(t) = \frac{1}{2}$ .

We next decompose the nonequilibrium density  $\rho(x, t)$ , using the law of total probability, into left and right components:

$$\rho(x, t) = p(t) \rho(x, t|x < 0) + [1 - p(t)] \rho(x, t|x > 0). \quad (9)$$

In contrast to the form given in Eq. 8, the nonequilibrium  $\rho$  makes no hypotheses as to the form of the conditional densities. However, the function  $p$  is chosen to be the same in both densities. Because Eq. 9 simply applies the definition of conditional probabilities, it is always possible to write the nonequilibrium density in this way.

Interpreting the entropy  $S$  as the Gibbs-Shannon entropy associated to the nonequilibrium density, Eq. 1, the nonequilibrium free energy  $F_{\text{neq}}$  can be expressed in terms of the local equilibrium as

$$F_{\text{neq}}(t) = F_{\text{leq}}(t) - k_B T (\ln 2) H[p(t)] + k_B T D_{\text{KL}}[\rho(x, t) || \rho_{\text{leq}}(x, t)], \quad (10)$$

where  $H[p(t)]$  is the discrete binary Shannon entropy (in bits),

$$H(p) = -p \log_2 p - (1 - p) \log_2 (1 - p), \quad (11)$$

and where the relative entropy (Kullback-Leibler divergence) is [18]

$$D_{\text{KL}}[p(x) || q(x)] \equiv \int_{-\infty}^{\infty} dx p(x) \ln \left( \frac{p(x)}{q(x)} \right), \quad (12)$$

for probability density functions  $p(x)$  and  $q(x)$ . Equation 10 can easily be generalized to an asymmetric multi-well potential; particular cases are proved in [44, 46, 47]. Note that for  $0 < p < 1$ , the Shannon entropy  $H(p)$  ranges between 0 and 1 bit, and the relative entropy measures the distinguishability of two probability distributions and satisfies  $D_{\text{KL}}[p(x) || q(x)] \geq 0$ , equaling zero

only when  $p(x) = q(x)$ . See the Supplement for a derivation of Eq. 10. The second law with discrete entropy is then found by combining Eqs. 5 and 10. Note that the relative-entropy term quantifies the effect of the departure from local equilibrium in the second law, an issue that has been studied from a different point of view in Ref. [48].

## PROTOCOLS FOR MEASURING THE FUNCTION $H(p)$

The main idea is that, for slow, thermodynamically reversible protocols, the inequality in Eq. 5 becomes an equality, giving with Eq. 10 a way to obtain the function  $H(p)$  experimentally. To isolate the discrete entropy, we consider first a cyclic protocol that starts and ends with the system having the same symmetric double-well potential  $U(x)$ . This eliminates the free-energy difference  $\Delta F_{\text{leq}}$ . Moreover, we choose the initial density to always be in local equilibrium, and we choose protocol times  $\tau$  that are large enough that the final protocol is in local equilibrium, too, in the potential  $U(x)$ . (Of course, here and elsewhere in this paper, we always assume that the protocol time  $\tau$  is shorter than the time to globally equilibrate via spontaneous hops over the barrier; that time scale is effectively infinite.) The relative entropy term in Eq. 10 then vanishes at both  $t = 0$  and  $t = \tau$ . Finally, under these conditions, the change in non-equilibrium free energy is simply, from Eq. 10,

$$\Delta F_{\text{neq}} = -k_B T (\ln 2) \Delta H, \quad (13)$$

This is the principle proposed by Landauer in 1961 [35] and studied extensively since [41, 42, 44, 46–59], with recent experimental confirmation [60–64]. Thus, by measuring the minimal average work to carry out protocols that alter the information content of a two-state system, we can test whether the Shannon entropy has physical relevance: Does it apply to thermodynamic descriptions such as Eq. 5?

More precisely, we explore experimentally the two protocols illustrated in Fig. 1:

- *Protocol 1:* We erase completely a fraction of a bit of information. The initial state of the system is a local equilibrium, with a probability  $p_0$  for a particle to be in the left well. The state encodes an information content  $H_0 = H(p_0)$ . At the end of the protocol, at time  $\tau$ , the particle is again in local equilibrium but now always in the right well, implying that  $H_\tau = 0$ . Thus,  $\Delta H = -H_0$  and  $\Delta F_{\text{neq}} = k_B T (\ln 2) H_0$ .
- *Protocol 2:* We start with one bit of information and erase a fraction of it. The initial state of the system is local equilibrium with  $p_0 = \frac{1}{2}$ , which corresponds to one bit of information. The final state,

after time  $\tau$ , is in local equilibrium with probability  $p_\tau$  to be in the left well, corresponding to  $H_\tau$  between zero and one bit. Thus,  $\Delta H = H_\tau - 1$  and  $\Delta F_{\text{neq}} = k_B T (\ln 2) [1 - H_\tau]$ .

This protocol resembles that used in [60, 65, 66]. However, in those studies, partial erasure was used because the barrier could not be made high enough to ensure full erasure, and correction factors were applied to infer the work required for full erasure of a bit. Here, we will use, in a controlled way, the partial work as a means to estimate the Shannon entropy function,  $H(p)$ .

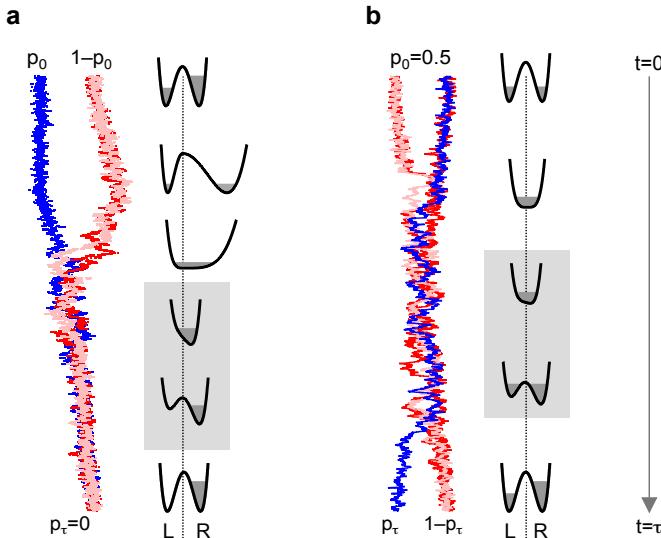


FIG. 1. Protocols of duration  $\tau$  for erasing a fraction of a bit, accompanied by sample trajectories. (A) Protocol 1: full erasure of a fractional bit. The potential is stretched to bring the two states to global equilibrium before mixing. Full erasure is achieved using a strong tilt (gray shading). One trajectory (blue) starts in the left well; two (red, pink) start in the right. All end in the right well. (B) Protocol 2: fractional erasure of a full bit. Initial equilibrium state is mixed directly. Weak tilt (gray shading) controls the final probability. A quarter of the trajectories end in the left well.

Figure 1a shows Protocol 1. Naively, one might lower the barrier as a first step; however, such a protocol leads experimentally (and analytically) to an asymptotic work of  $k_B T \ln 2$  for all initial probabilities  $p_0$  (see supplement). But first stretching the potential to bring the system to global equilibrium before lowering the barrier allows it to reach the reversible bound,  $k_B T (\ln 2) H(p_0)$ . We thus stretch, lower the barrier, compress, strongly tilt, raise the barrier, and finally untilt to return the potential to its initial shape. For a strong tilt, all observed trajectories end in the right well.

## RESULTS

For each cycle time  $\tau$  and each initial state, we find the average work. Figure 2a shows the average conditional work for particles starting in the left and right wells. Figure 2b shows the combined average work. Work in the slow limit is estimated by extrapolating to long times. In this limit, the protocol is fully reversible, and the nonequilibrium free-energy change equals the work done by the potential,  $\Delta F_{\text{neq}} = W_\infty$ . We plot the scaled change in nonequilibrium free energy  $\Delta F_{\text{neq}}/k_B T$  as a function of  $p_0$  in Fig. 3a.

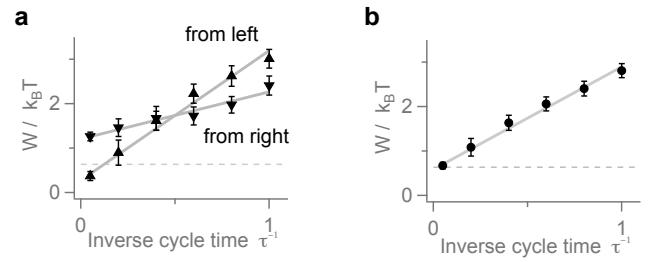


FIG. 2. Work to erase a fraction of a bit (Protocol 1). (A) Conditional work measurements for particles starting in the left and right wells. (B) Unconditioned work required to erase a fraction of a bit for  $p_0 = \frac{1}{3}$  at finite times  $\tau$ . Extrapolating the fit gives  $W_\infty/k_B T = 0.58 \pm 0.07$ , with  $\chi^2 = 1.4$  for 4 degrees of freedom. The dashed horizontal lines denote  $(\ln 2)$  times the change in information in bits:  $(\ln 2) \Delta H = (\ln 2) H(\frac{1}{3}) \approx 0.64$ , as calculated from Eq. 11.

Our measurements show that it takes less than  $k_B T \ln 2$  of work to erase less than one bit of information. Although the results from Protocol 1 are consistent with the expected shape of the Shannon entropy function,  $(\ln 2) H(p_0)$ , they test only a narrow range of  $p_0$ , since large stretching factors  $\eta$  imply long time scales ( $\sim \eta^2$  because of diffusion).

To explore a wider range of information erasure, we therefore developed a second protocol that tilts rather than stretches the potential to create an energy difference between two local minima. Tilting a potential does not increase its spatial extent and allows us to explore the full change of information from 0 to 1 bit. However, there are problems that preclude extrapolating small-tilt protocols to long times (see supplement).

We thus designed a protocol that operates at a *fixed*, large cycle time  $\tau$ . At fixed  $\tau$ , the mean work  $W(\tau)$  needed to change the information from  $H_0$  to  $H_\tau$  is always strictly greater than the change in free energy  $W > \Delta F_{\text{neq}}$  (Fig. 2b). To isolate the lower bound of the work, we run the protocol in the forward and then the backward direction. When the protocol is executed slowly enough that conditional work distributions are

Gaussian, we find (see supplement)

$$\frac{1}{2} (W_F - W_B) = \Delta F_{\text{neq}} = k_B T (\ln 2) [1 - H_\tau], \quad (14)$$

where  $1 - H_\tau$  is minus the change in Shannon entropy and  $W_F$  ( $W_B$ ) the average work for the forward (backward) part of the protocol. Similar formulas have been used to estimate *equilibrium* free energy differences [67, 68]. Here, we estimate the *nonequilibrium* free energy difference using Eq. 14.

Figure 3a shows the results of Protocol 2 (hollow markers), plotted as  $\ln 2 - \Delta F_{\text{neq}}/k_B T$  so that the data from Protocols 1 and 2 may be compared directly. The plot agrees—without fit—with the Gibbs-Shannon form,  $(\ln 2) H(p)$ , over the full range  $p \in [0, 1]$ . Figure 3b then shows that this change in nonequilibrium free energy is linear in the Shannon entropy change.

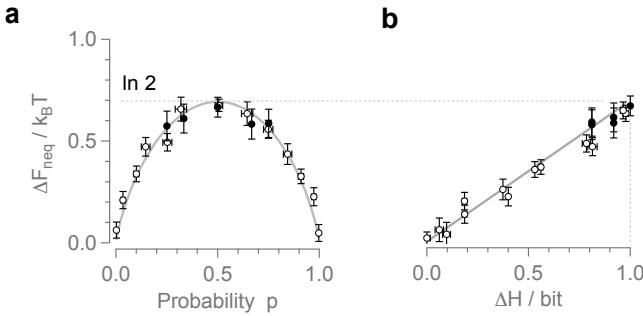


FIG. 3. Change in non-equilibrium free energy due to a partial memory erasure. Filled markers are measured using Protocol 1 by extrapolation, hollow markers using Protocol 2 at fixed cycle time  $\tau = 2$ . (A) Plot vs. probability,  $p_0$  and  $p_\tau$ , respectively, in the two protocols. Solid gray line is a plot of  $H(p)$ , with no fit parameters. (B) Plot vs. change in Shannon entropy, in the limit of slow protocols. Solid line—not a fit—shows the predicted slope of  $\ln 2 \approx 0.69$  per bit, from Eq. 5.

## DISCUSSION

Although our focus in this paper is on testing the Gibbs-Shannon entropy for discrete states, Eq. 11, we measure a continuous position and can test explicitly aspects of the continuum version of the entropy, Eq. 1. For example, we argue in the supplement that our data are consistent with a conditioned version of Crooks' relation. Further, we discuss how the measurements presented here also confirm the identification between the total entropy production  $S_{\text{tot}}$  and the relative entropy between the forward and backward path measures.

Beyond their role in justifying the underlying assumptions made in the field of stochastic thermodynamics, our results may aid continuing efforts to understand the

role of information in nonequilibrium biological systems, where cells actively sense their environment and respond. For example, we saw that a naive version of Protocol 1 was intrinsically irreversible and therefore unable to reach the ultimate thermodynamic bounds based on starting and ending states. In recently published work [69], Ouldridge et al. argue that realistic biochemical networks similarly cannot reach these fundamental bounds. In that case, the authors trace the extra dissipation to a failure to exploit all correlations generated between the measuring device and the physical system (receptors and readouts). It will be interesting to study systematically the various classes of explanations for dissipation beyond the minimum levels reached here in a more-idealized kind of experiment.

## CONCLUSION

Two different protocols that each measure the minimal average work required to erase a fraction of a bit of information both confirm that the nonequilibrium system entropy of a colloidal particle in a controllable potential has a functional form consistent with that proposed long ago by Gibbs and Shannon.

## Experimental Setup

A feedback (or Anti-Brownian EElectrokinetic, or ABEL) trap is a technique for trapping and manipulating small particles in solution [70]. The basic idea is to replace a trapping potential with a feedback loop: in one cycle, one measures the position of a particle and then applies a force (created by an electric field) that pushes it back to the desired trapping point. By the next cycle, thermal fluctuations have pushed the particle in a different direction, and a new restoring force is computed. Feedback traps can also be used to place particles in a virtual potential, where the motion imitates a desired potential [31, 61, 71, 72].

In the protocols described below, we take advantage of the nearly complete freedom to specify arbitrarily the shape of a virtual potential. Thus, we can selectively lower the barrier, while keeping the outer part of the potential fixed. Or, we can selectively stretch one well by a factor  $\eta$  while the other well is unchanged. Such manipulations are not possible in erasure experiments based on optical tweezers [60, 65, 66], which limits the possible protocols in such cases.

The challenge with using feedback traps to measure work values to an accuracy  $< 0.1 k_B T$  is to calibrate forces accurately and to account for slow drifts in quantities such as the particle's response to an applied voltage. In earlier work, we developed a recursive, real-time calibration technique [73] that allows us to measure accurately the stochastic work done by a changing potential

on a particle. Using an improved setup with higher feedback loop rates [74], we explored erasure in asymmetric memories [75], tested subtle forms of reversibility [76], and compared different estimators of heat transfer [77].

The experimental setup for our feedback trap has three major segments: the imaging system, the trapping chamber, and the control software. The imaging system consists of an inverted, home-built, dark-field, front-illumination microscope with a 60x Olympus NA=0.95 air objective [74, 78]. A silica bead of diameter  $1.5 \mu\text{m}$  is illuminated by a 660-nm LED source. A small disk placed behind the objective blocks the direct LED light but allows scattered light to reach a camera. The camera (Andor iXon DV-885) takes a  $50 \times 20$  pixel image every  $\Delta t = 5$  ms, with an exposure  $t_c = 0.5$  ms. The trapping chamber is cylindrical,  $\approx 10$  mm in diameter and  $\approx 5$  mm in height, and is glued on top of a glass coverslip. We load silica beads diluted in deionized water. The beads sink to the bottom of the chamber (top of the coverslip) under gravity, which confines them in the vertical ( $z$ ) direction. Two pairs of electrodes near the bottom of the chamber create an electric field ( $\sim 10$  V/cm) whose value is updated every time step to move a bead in the horizontal ( $xy$ ) plane [75–77]. The control software analyzes images in real time using a centroid algorithm [79]. It calculates forces based on the measured position and value of the gradient of the virtual potential. Simultaneously, deviations between the expected and measured positions are used to calibrate the feedback trap, using a recursive maximum likelihood algorithm for a continuous linear fit between the applied voltages and observed displacements [73]. The particle's electric-field mobility is estimated from the slope. Drifts are assessed via the intercept, and particle diffusion is estimated from the fit residuals. A running-average algorithm keeps only the most recent measurements and helps track parameter changes during experiments that can last several days.

Finally, in the supplement, we justify in more detail our model of the dynamics as one dimensional and over-damped [80].

### Data Analysis

Feedback traps allow one to impose an arbitrary virtual potential of almost any form. We choose a static harmonic potential in the  $y$  direction and a double-well potential in the  $x$  direction:

$$U(x, t) = 4E_b \left[ -\frac{1}{2}g(t)\tilde{x}^2 + \frac{1}{4}\tilde{x}^4 - Af(t)\tilde{x} \right], \quad (15)$$

where the scaled coordinate  $\tilde{x}(x, t)$  is selectively stretched for positive or negative  $x$ , as desired. More precisely,  $\tilde{x}(x < 0, t) = -\tilde{\eta}(t)\tilde{x}(x \geq 0, t)$  where  $\tilde{\eta}(t)$  is a time-dependent stretching factor (see Fig. 1a). Note that the stretching amplitude  $\eta$  scales the stretching factor  $\tilde{\eta}(t)$ . In all cases,  $\tilde{\eta}(0) = \tilde{\eta}(\tau) = 1$ , so that we start and end

with a symmetric potential. In Eq. 15,  $E_b$  is the energy barrier height,  $A$  the tilt amplitude. The functions  $g(t)$  and  $f(t)$  can take values between 0 and 1 and control the barrier height and tilt. Together with stretching  $\tilde{\eta}(t)$ , they allow us to implement Protocols 1 and 2, as described below and in the Supplement.

Each experiment uses several beads, whose properties must each be measured using the recursive algorithm given above. Via dimensionless scaling, we can combine data measured on beads, which, although nominally identical, differ slightly in radius and charge. The measured diffusion constant near the surface is typically  $\approx 0.23 \mu\text{m}^2/\text{s}$ . Based on the requirement that feedback update time be much smaller than the local relaxation time within a well, we set the distance between two local minima of the double-well potential. A typical value is  $2x_0 = 1.54 \mu\text{m}$ . The dimensionless time  $\tau = 1$  then corresponds to a physical time  $t_{\text{sec}} \approx 10$  s.

The work to manipulate a potential in one cycle of duration  $\tau$  is estimated by discretizing Sekimoto's formula [29, 40] for the stochastic work,

$$w_\tau = \int_0^\tau dt \left. \frac{\partial U(x, t)}{\partial t} \right|_{x=x(t)}. \quad (16)$$

### Experimental Protocols

We used two different erasure protocols. In both, we prepare the initial state by placing a particle in a given well using a strong harmonic trap for 0.5 s. We then abruptly switch to a static double-well potential to let a particle equilibrate locally for 1 s, before the cycle starts. Below, we describe qualitatively each protocol. (See Supplement for the explicit potentials,  $U(x, t)$ .)

#### Protocol 1

The initial state is in the left well with probability  $p_0$  and has system entropy  $H_0 = H(p_0)$ . We erase to a state with  $p_\tau = 0$  (always in the right well) and  $H_\tau = 0$ . We define the initial state of the memory by placing a particle in a particular well. The high energy barrier of  $E_b = 13 k_B T$  prevents the two states from mixing on the time scales of the experiment.

We measure the mean work for full erasure from this initial state via conditional work values. That is, we measure the average value of work  $W_L$  to erase conditioned on starting in the left well and similarly for the right well,  $W_R$ . For  $N_L$  individual measurements  $w_L^i$ , we estimate the mean via the average,  $W_L \approx \bar{W}_L = \frac{1}{N_L} \sum_i w_L^i$ . Similarly,  $W_R \approx \bar{W}_R = \frac{1}{N_R} \sum_i w_R^i$ . The unconditional work at time  $\tau$  is estimated from the law of total probability as  $W_\tau = p_0 W_L + (1 - p_0) W_R$ . The work in the slow limit

is obtained by extrapolating using the asymptotic form  $W_\tau \sim W_\infty + a\tau^{-1}$  and fitting a line against  $\tau^{-1}$  [57, 81].

We need to start by stretching the potential by a factor  $\eta = 1/p_0 - 1$ , to equalize the probability densities in the left and right states and bring them to global equilibrium. Otherwise, lowering the barrier would be an irreversible step that adds dissipation that does not vanish, even in the slow limit [76].

For  $p_0 \geq 0.5$ , the left well is stretched, while, for  $p_0 < 0.5$ , the right well is stretched. (At  $\eta = 1$ , the wells have their minimum width, a width set by requiring that gradients be small enough that the discrete approximation to a continuous potential is accurate [71]. We thus stretch one or the other well, depending on  $p_0$ .) Note that, as a consequence of the stretching, the values of  $W_L$  and  $W_R$  depend on  $p_0$ . After stretching, we lower the barrier and mix the states, then strongly tilt towards the right. Finally, we increase the barrier and untilt the potential. This protocol is repeated for several different cycle times  $\tau$ , where, for each  $\tau$ , we recorded multiple trajectories over a thirty-minute period. The uncertainty in the estimate of average work values depends only on the total time of data collection, not on the cycle time  $\tau$  directly [77].

#### Protocol 2

The initial state has one bit of information, which is erased partially. The initial state at time  $t = 0$  is in global equilibrium, with  $p_0 = 0.5$  and  $H_0 = 1$  bit, and ends with  $H_\tau$ , which we control in the range from 0 to 1 bit. The slightly lower energy barrier  $E_b = 10 k_B T$  reduces the distance between wells, which must be large enough that the virtual potential lead to dynamics that are indistinguishable from those of the corresponding physical potential [71]. Because the fixed cycle time is short ( $\approx 30$  s), the probability of a spontaneous hop over the barrier is negligible.

Protocol 2 operates at the fixed cycle time  $\tau = 2$ . In four steps, we lower the barrier and mix states, apply a weak tilt with an amplitude  $A$ , raise the barrier, and untilt. The entire protocol is then repeated in reverse. For each tilt  $A$ , we acquire data for about 12 hours. We measure the stochastic work from each trajectory and the probability to end in the left well  $p_\tau$  after the forward protocol. As a control, we estimate the probability to end in the left well after reverse protocol, which is consistent with the expected value of 0.5 for a reversible protocol. (See supplemental material for data.)

Ensemble averages for Protocol 2 are estimated from the arithmetic mean of  $N$  work measurements in the forward and reverse protocols:  $W_F \approx \bar{W}_F = \frac{1}{N} \sum_i W_F^i$  and  $W_B \approx \bar{W}_B = \frac{1}{N} \sum_i W_B^i$ . By recording the work done for forward and backwards protocols at a fixed cycle time  $\tau$ , we have a simple, accurate way to estimate the change in

nonequilibrium free energy (see section 4 of supplement). Error bars on work measurements in all cases represent the standard error of mean, calculated as  $\sigma_W / \sqrt{N}$ , with  $\sigma_W$  the standard deviation of the  $N$  individual measurements.

#### ACKNOWLEDGMENT

We thank J. Vollmer, D. Sivak, É. Roldán, H. Touchette, and M. Esposito for suggestions. This work was supported by a Discovery Grant from NSERC (Canada). R.C. was supported by the French Ministry of Education through grant ANR-15-CE40-0020-01.

- 
- [1] H. Grad, “The many faces of entropy,” Comm. Pure and Appl. Math. **14**, 323–354 (1961).
  - [2] O. Penrose, *Foundations of Statistical Mechanics: A Deductive Treatment* (Pergamon, 1970).
  - [3] A. Wehrl, “General properties of entropy,” Rev. Mod. Phys. **50**, 221–260 (1978).
  - [4] M. C. Mackey, “The dynamical origin of increasing entropy,” Rev. Mod. Phys. **61**, 981–1015 (1989).
  - [5] A. Wehrl, “The many facets of entropy,” Rep. Math. Phys. **30**, 119–129 (1991).
  - [6] R. Balian, “Entropy, a protean concept,” Sem. Poincaré **2**, 13–27 (2003).
  - [7] E. T. Jaynes, “E. t. jaynes: Papers on probability, statistics and statistical physics,” (Dordrecht, Holland, 1983).
  - [8] J. Bricmont, “Science of chaos or chaos in science,” Ann. (N.Y.) Acad. Sci. **79**, 131–175 (1996).
  - [9] J. Dunkel and S. Hilbert, “Consistent thermostatistics forbids negative absolute temperatures,” Nat. Phys. **10**, 67–72 (2014).
  - [10] J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Yale University Press, 1902).
  - [11] C. E. Shannon, “A mathematical theory of communication,” Bell Syst. Tech. J. **27**, 379–423, 623–656 (1948).
  - [12] J. von Neumann, “Thermodynamik quantummechanischer Gesamtheiten,” Gött. Nach. **1**, 273–291 (1927).
  - [13] B.-S. K. Skagerstam, “On the notions of entropy and information,” J. Stat. Phys. **12**, 449–462 (1975).
  - [14] J. L. Lebowitz, “Boltzmann’s entropy and time’s arrow,” Phys. Today **46**, 32–38 (1993).
  - [15] S. Goldstein and J. L. Lebowitz, “On the (Boltzmann) entropy of non-equilibrium systems,” Physica D **193**, 53–66 (2004).
  - [16] M. Hemmo and O. R. Shenker, *The Road to Maxwell’s Demon: Conceptual Foundations of Statistical Mechanics* (Cambridge Univ. Press, 2012).
  - [17] O. C. O. Dahlsten, R. Renner, E. Rieper, and V. Vedral, “Inadequacy of von Neumann entropy for characterizing extractable work,” New J. Phys. **13**, 053015 (2011).
  - [18] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. (John Wiley & Sons, Inc., New York, 2006).
  - [19] E. H. Lieb and J. Yngvason, “The entropy concept for non-equilibrium states,” Proc. R. Soc. A **469**, 20130408 (2013).
  - [20] M. Weilenmann, L. Kraemer, P. Faist, and R. Ren-

- ner, "Axiomatic relation between thermodynamic and information-theoretic entropies," *Phys. Rev. Lett.* **117**, 260601 (2016).
- [21] C. Jarzynski, "Nonequilibrium equality for free energy differences," *Phys. Rev. Lett.* **78**, 2690–2693 (1997).
- [22] G. E. Crooks, "Non-equilibrium measurements of free energy differences for microscopically reversible Markovian systems," *J. Stat. Phys.* **90**, 1481–1487 (1998).
- [23] J. L. Lebowitz and H. Spohn, "A Gallavotti-Cohen type symmetry in the large deviation functional for stochastic dynamics," *J. Stat. Phys.* **95**, 333–365 (1999).
- [24] C. Maes, "The fluctuation theorem as a Gibbs property," *J. Stat. Phys.* **95**, 367–392 (1999).
- [25] C. Maes, K. Netočný, and B. Shergelashvili, "A selection of nonequilibrium issues," in *Methods of Contemporary Mathematical Statistical Physics*, edited by R. Kotecký (Springer, 2009) pp. 247–306.
- [26] C. Jarzynski, "Hamiltonian derivation of a detailed fluctuation theorem," *J. Stat. Phys.* **98**, 77–102 (2000).
- [27] U. Seifert, "Entropy production along a stochastic trajectory and an integral fluctuation theorem," *Phys. Rev. Lett.* **95**, 040602 (2005).
- [28] R. Chétrite and K. Gawędzki, "Fluctuation relations for diffusion processes," *Commun. Math. Phys.* **282**, 469–518 (2008).
- [29] K. Sekimoto, *Stochastic Energetics* (Springer, 2010).
- [30] U. Seifert, "Stochastic thermodynamics, fluctuation theorems and molecular machines," *Rep. Prog. Phys.* **75**, 1–58 (2012).
- [31] A. E. Cohen, "Control of nanoparticles with arbitrary two-dimensional force fields," *Phys. Rev. Lett.* **94**, 118102 (2005).
- [32] C. W. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 4th ed. (Springer, 2009).
- [33] H. Risken, *The Fokker Planck Equation*., 2nd ed. (Springer, Berlin-Heidelberg, 1989).
- [34] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed., North-Holland Personal Library (North Holland, 2007).
- [35] R. Landauer, "Irreversibility and heat generation in the computing process," *IBM J. Res. Develop.* **5**, 183–191 (1961).
- [36] D. Kondepudi and I. Prigogine, *Modern Thermodynamics: From Heat Engines to Dissipative Structures* (Wiley, 2014).
- [37] S. R. de Groot and P. Mazur, *Non-equilibrium Thermodynamics* (North Holland Pub. Co., 1962).
- [38] C. Maes and K. Netočný, "Time-reversal and entropy," *J. Stat. Phys.* **110**, 269–310 (2003).
- [39] R. Clausius, "Ueber die bewegende Kraft der Wärme und die Gesetze, welche sich daraus für die Wärmelehre selbst ableiten lassen. II," *Annalen der Physik* **79**, 500–524 (1850).
- [40] K. Sekimoto, "Kinetic characterization of heat bath and the energetics of thermal ratchet models," *J. Phys. Soc. Jap.* **66**, 1234–1237 (1997).
- [41] M. Esposito and C. Van den Broeck, "Second law and Landauer principle far from equilibrium," *EPL (Europhysics Letters)* **95**, 40004 (2011).
- [42] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, "Thermodynamics of information," *Nature Phys.* **11**, 131–139 (2015).
- [43] I. Junier, A. Mossa, M. Manosas, and F. Ritort, "Recovery of free energy branches in single molecule experiments," *Phys. Rev. Lett.* **102**, 070602 (2009).
- [44] T. Sagawa, "Thermodynamic and logical reversibilities revisited," *J. Stat. Mech.* , P03025 (2014).
- [45] É. Roldán, I. A. Martínez, J. M. R. Parrondo, and D. Petrov, "Universal features in the energetics of symmetry breaking," *Nature Phys.* **10**, 457–461 (2014).
- [46] K. Shizume, "Heat generation required by information erasure," *Phys. Rev. E* **52**, 3495–3499 (1995).
- [47] T. Sagawa and M. Ueda, "Minimal energy cost for thermodynamic information processing: Measurement and information erasure," *Phys. Rev. Lett.* **102**, 250602 (2009).
- [48] D. Chiuchiú, M. C. Diamantini, and L. Gamaitoni, "Conditional entropy and Landauer principle," *EPL (Europhysics Letters)* **111**, 40004 (2015).
- [49] C. H. Bennett, "The thermodynamics of computation: a review," *Int. J. Theor. Phys.* **21**, 905–940 (1982).
- [50] R. Landauer, "Information is physical," *Phys. Today* **44**, 23–29 (1991).
- [51] B. Piechocinska, "Information erasure," *Phys. Rev. A* **61**, 062314 (2000).
- [52] M. B. Plenio and V. Vitelli, "The physics of forgetting: Landauer's erasure principle and information theory," *Contemp. Phys.* **42**, 25–60 (2001).
- [53] H. S. Leff and A. F. Rex, *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing* (IOP, 2003).
- [54] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, "Dissipation: The phase-space perspective," *Phys. Rev. Lett.* **98**, 080602 (2007).
- [55] R. Dillenschneider and E. Lutz, "Memory erasure in small systems," *Phys. Rev. Lett.* **102**, 210601 (2009).
- [56] Koji Maruyama, Franco Nori, and Vlatko Vedral, "The physics of Maxwell's demon and information," *Rev. Mod. Phys.* **81**, 1–23 (2009).
- [57] E. Aurell, K. Gawędzki, C. Mejía-Monasterio, R. Mohtayee, and P. Muratore-Ginanneschi, "Refined second law of thermodynamics for fast random processes," *J. Stat. Phys.* **147**, 487–505 (2012).
- [58] V. Jakšić and C.-A. Pillet, "A note on the Landauer principle in quantum statistical mechanics," *J. Math. Phys.* **55**, 075210 (2014).
- [59] E. Lutz and S. Ciliberto, "Information: From Maxwell's demon to Landauer's eraser," *Phys. Today* **68**, 30–35 (2015).
- [60] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, "Experimental verification of Landauer's principle linking information and thermodynamics," *Nature* **483**, 187–189 (2012).
- [61] Y. Jun, M. Gavrilov, and J. Bechhoefer, "High-precision test of Landauer's principle in a feedback trap," *Phys. Rev. Lett.* **113**, 190601 (2014).
- [62] J. Hong, B. Lambson, S. Dhuey, and J. Bokor, "Experimental test of Landauer's principle in single-bit operations on nanomagnetic memory bits," *Sci. Adv.* **2**, e1501492 (2016).
- [63] L. Martini, M. Pancaldi, M. Madami, P. Vavassori, G. Gubbiotti, S. Tacchi, F. Hartmann, M. Emmerling, S. Höfling, L. Worschech, and G. Carlotti, "Experimental and theoretical analysis of Landauer erasure in nanomagnetic switches of different sizes," *Nano Energy* **19**, 108–116 (2016).
- [64] J. P. S. Peterson, R. S. Sarthour, A. M. Souza, I. S.

- Oliveira, J. Goold, K. Modi, D. O. Soares-Pinto, and L. C. Céleri, “Experimental demonstration of information to energy conversion in a quantum system at the Landauer limit,” Proc. R. Soc. A **472**, 20150813 (2016).
- [65] A. Bérut, A. Petrosyan, and S. Ciliberto, “Detailed Jarzynski equality applied to a logically irreversible procedure,” EPL **103**, 60002 (2013).
- [66] A. Bérut, A. Petrosyan, and S. Ciliberto, “Information and thermodynamics: experimental verification of Landauer’s Erasure principle,” J. Stat. Mech. , P06015 (2015).
- [67] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante, “Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies,” Nature **437**, 231–234 (2005).
- [68] S. Kim, Y. W. Kim, P. Talkner, and J. Yi, “Comparison of free-energy estimators and their dependence on dissipated work,” Phys. Rev. E **86**, 041130 (2012).
- [69] T. E. Ouldridge, C. C. Govern, and P. R. ten Wolde, “Thermodynamics of computational copying in biochemical systems,” Phys. Rev. X **7**, 021004 (2017).
- [70] A. E. Cohen and W. E. Moerner, “Method for trapping and manipulating nanoscale objects in solution,” App. Phys. Lett. **86**, 093109 (2005).
- [71] Y. Jun and J. Bechhoefer, “Virtual potentials for feedback traps,” Phys. Rev. E **86**, 061106 (2012).
- [72] M. Gavrilov, Y. Jun, and J. Bechhoefer, “Particle dynamics in a virtual harmonic potential,” Proc. SPIE **8810** (2013).
- [73] M. Gavrilov, Y. Jun, and J. Bechhoefer, “Real-time calibration of a feedback trap,” Rev. Sci. Instrum. **85**, 095102 (2014).
- [74] M. Gavrilov, J. Koloczek, and J. Bechhoefer, “Feedback trap with scattering-based illumination,” in *Novel Techniques in Microscopy* (Opt. Soc. Am., 2015) p. JT3A. 4.
- [75] M. Gavrilov and J. Bechhoefer, “Erasure without work in an asymmetric double-well potential,” Phys. Rev. Lett. **117**, 200601 (2016).
- [76] M. Gavrilov and J. Bechhoefer, “Arbitrarily slow, non-quasistatic, isothermal transformations,” EPL (Euro-physics Letters) **114**, 50002 (2016).
- [77] M. Gavrilov and J. Bechhoefer, “Feedback traps for virtual potentials,” Phil. Trans. R. Soc. A **375**, 20160217 (2017).
- [78] A. Weigel, A. Sebesta, and P. Kukura, “Dark field microspectroscopy with single molecule fluorescence sensitivity,” ACS Photonics **1**, 848–856 (2014).
- [79] A. J. Berglund, M. D. McMahon, J. J. McClelland, and J. A. Liddle, “Fast, bias-free algorithm for tracking single particles with variable size and shape,” Opt. Express **16**, 14064–14075 (2008).
- [80] J. Happel and H. Brenner, *Low Reynolds Number Hydrodynamics: With Special Applications to Particulate Media* (Martinus Nijhoff, 1983).
- [81] K. Sekimoto and S. Sasa, “Complementarity relation for irreversible process derived from stochastic energetics,” J. Phys. Soc. Jap. **66**, 3326–3328 (1997).

**Supplementary Information:**  
**Direct measurement of nonequilibrium system entropy is consistent with**  
**Gibbs-Shannon form**

Momčilo Gavrilov<sup>1,†</sup>, Raphaël Chérite<sup>1,2,3</sup>, and John Bechhoefer<sup>1\*</sup>

<sup>1</sup> Department of Physics, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada

<sup>2</sup> Pacific Institute for the Mathematical Sciences,  
UMI 3069, Vancouver, British Columbia, Canada

<sup>3</sup> Université Côte d'Azur, CNRS, LJAD, Parc Valrose, 06108 NICE Cedex 02, France

†Present address: Department of Biophysics and Biophysical Chemistry,  
Johns Hopkins University, 725 N. Wolfe Street, Baltimore, MD 21205-2185, USA

## I. SYSTEM DYNAMICS

The system dynamics are described as one dimensional and overdamped. Here, we give a brief justification for this claim. The inertial damping time of a micron-scale bead in water  $\sim \Delta m/\gamma \approx 10^{-6}$  s, where  $\Delta m$  is the difference in mass between the bead and the fluid it displaces and where  $\gamma$  is the drag coefficient. This time is much shorter than the shortest time scale probed in the experiment, the position-measurement time due to the camera exposure,  $\Delta t = 2 \cdot 10^{-4}$  s, and can thus be ignored. The equation of motion is thus a one-dimensional, overdamped Langevin equation of the form given in Eq. 2. Note that the friction coefficient  $\gamma$  given there can be calculated from hydrodynamics. For example, for a sphere of radius  $a$  moving in an unbounded fluid of viscosity  $\eta$ , the Stokes solution [1] describing hydrodynamic flow around the sphere implies  $\gamma = 6\pi\eta a$ . For our case, a sphere near a surface, extra drag due to the surface in-

creases  $\gamma$  [1]. The increase in  $\gamma$  is apparent through a reduction of the diffusion coefficient,  $D = k_B T / \gamma$ . In our case, we observed  $D/D_\infty \approx 0.67$ , where  $D_\infty$  is the value of the diffusion constant predicted using the Stokes drag expression.

Although the particle moves in three-dimensional space, it is confined in two of the dimensions,  $y$  and  $z$ . The motion in  $y$  is confined by imposing a virtual potential  $U_y(y) \sim \frac{1}{2}k_y y^2$  for deviations from the desired  $y$  position. The motion in  $z$  is confined by a physical potential that is a balance of electrostatic repulsion between the silica bead and the glass surface and the gravitational attraction. The size and density of the bead are chosen so that the bead is “slightly heavy”: it sinks to the bottom but nonetheless fluctuates about an equilibrium height  $\approx 0.2$  μm above the glass substrate. Since the  $y$  and  $z$  dependence of the potential is static, those variables play no role in the thermodynamics, and we can regard the virtual potential as an effectively one-dimensional potential,  $U(x, t)$ .

---

\* email: johnb@sfu.ca

## II. DERIVATION OF EQ. 10

Starting from  $F_{\text{neq}}(t) = E(t) - TS(t)$  and writing, as usual,  $\beta = (k_B T)^{-1}$  to simplify the notation, we have,

$$\begin{aligned}
 & (k_B)^{-1} S(t) + D_{\text{KL}}(\rho(x, t) \parallel \rho_{\text{leq}}(x, t)) \\
 &= - \int_{-\infty}^{\infty} dx \rho(x, t) \ln \rho(x, t) + \int_{-\infty}^{\infty} dx \rho(x, t) \log \left( \frac{\rho(x, t)}{\rho_{\text{leq}}(x, t)} \right) \\
 &= - \int_{-\infty}^{\infty} dx \rho(x, t) \ln \rho_{\text{leq}}(x, t) \\
 &= -p(t) \int_{-\infty}^0 dx \rho(x, t|x < 0) \ln [p(t) \exp[\beta(F_{\text{leq}} - U(x))]] \\
 &\quad - (1-p(t)) \int_0^{\infty} dx \rho(x, t|x > 0) \ln [(1-p(t)) \exp[\beta(F_{\text{leq}} - U(x))]] \\
 &= H[p(t)] - p(t) \int_{-\infty}^0 dx \rho(x, t|x < 0) [\beta(F_{\text{leq}} - U(x))] \\
 &\quad - (1-p(t)) \int_0^{\infty} dx \rho(x, t|x > 0) [\beta(F_{\text{leq}} - U(x))] \\
 &= H[p(t)] - p(t)\beta F_{\text{leq}} - (1-p(t))\beta F_{\text{leq}} \\
 &\quad + p(t)\beta E(t|x < 0) + (1-p(t))\beta E(t|x > 0) \\
 &= H[p(t)] - \beta[F_{\text{leq}} - E(t)]. \tag{S1}
 \end{aligned}$$

Thus,  $TS(t) + k_B T D_{\text{KL}}(\cdot \parallel \cdot) = k_B T H[p(t)] - F_{\text{leq}} + E(t)$ , and, finally,

$$\begin{aligned}
 F_{\text{neq}}(t) &= E(t) - TS(t) \\
 &= F_{\text{leq}} - k_B T H[p(t)] + k_B T D_{\text{KL}}(\rho(x, t) \parallel \rho_{\text{leq}}(x, t)), \tag{S2}
 \end{aligned}$$

which is Eq. 10.

Note that the local-equilibrium density function  $\rho_{\text{leq}}(x, t)$  is discontinuous at  $x = 0$  in general for  $p \neq \frac{1}{2}$ , as illustrated in Fig. S1. It will then be difficult to design a protocol that makes the  $D_{\text{KL}}$  term vanish at the end. However, if the barrier is high compared to  $k_B T$ , then  $\rho_{\text{leq}}(x) \approx 0$  in a finite interval about  $x = 0$ , allowing one to think of the density as approximately two independent conditional densities (Fig. S1a). In such a case, the one chosen for these experiments, it is possible to design a protocol that puts the system in local equilibrium with controllable  $p$  at the start and end. The  $D_{\text{KL}}$  terms then vanish in Eq. 11, making the isolation of the Shannon-entropy contribution to  $F_{\text{neq}}$  more direct and easier to extract.

## III. A NAIVE VERSION OF PROTOCOL 1

In Figure S2a, we illustrate a naive version of Protocol 1 that is a simple generalization of the protocol used to erase a full bit of information [2]. The system starts with probabilities  $p_0$  and  $1 - p_0$  for the particle to be in the left and right wells, respectively. At the end, the particle is always in the left well. Then, Eqs. (3) and (4) from the main text imply that the average work should be bounded

below by  $W \geq k_B T (\ln 2) H(p_0)$ . Naively, this suggests that, for sufficiently slow protocols, the asymptotic average work will be  $W = k_B T (\ln 2) H(p_0)$ . Instead, we measure  $\approx k_B T \ln 2$ , for all values of  $p_0$  (Figure S2b).

Intuitively, we can understand this result using Figure S2c. There, we plot the average work conditioned on whether the particle starts in the left (L) or right (R) wells,  $W_L$  or  $W_R$ . For large  $\tau$  (small  $\tau^{-1}$ ), the plots both converge to  $k_B T \ln 2$ . This makes sense: for slow protocols, the probability density has ample time to mix after the barrier is lowered. Thereafter, the two systems have the same density evolution. And as the barrier is lowered, the symmetry of the system ensures that the contributions to the work from particles starting in either state is the same. Thus, we must have  $W_L = W_R$ . But the average work for an initial state occupying the left well with probability  $p_0$  can be computed as follows:

$$\begin{aligned}
 W &= p_0 W_L + (1-p_0) W_R \\
 &= k_B T [p_0 (\ln 2) + (1-p_0) (\ln 2)] \\
 &= k_B T \ln 2. \tag{S3}
 \end{aligned}$$

Thus, two different lines of reasoning lead to two different lower bounds for the average work to erase. Moreover, the reasoning in both cases suggests that the bounds can be reached by extrapolating slow protocols to the large-

time limit, implying an incompatibility.

To resolve this contradiction, we note that when  $p_0 \neq 0.5$ , the system is not in global equilibrium. Lowering the barrier is then an irreversible step, because it allows the probabilities to mix. For example, reversing and raising the barrier (immediately after lowering) would lead to a state with probability 0.5 to be in each well, different from the initial state. Because of the dissipation associated with the irreversible mixing of probabilities, the average work must exceed the lower bound of  $k_B T (\ln 2) H(p_0)$  that is derived accounting only for the initial and final states. But why is the average work always  $k_B T \ln 2$ ?

To understand this last point more formally, we can use reasoning similar to that of Kawai et al. [3] to derive a refined version of the second law,

$$W \geq \Delta F_{\text{neq}} + k_B T D_{\text{KL}}[\rho_{\text{leq}}(x, 0) || \rho_{\text{eq}}(x)], \quad (\text{S4})$$

since the final density of the backward process is the global equilibrium state,  $\rho_{\text{eq}}(x)$ . The relative-entropy term captures the irreversibility of the protocol. An explicit calculation then gives  $D_{\text{KL}}[\rho_{\text{leq}}(x, 0) || \rho_{\text{eq}}(x)] = p_0 \ln(2p_0) + (1 - p_0) \ln[2(1 - p_0)] = \ln 2 - H(p_0)$ . Since  $\Delta F_{\text{neq}}/(k_B T) = \ln 2$ , we have

$$W \geq k_B T \ln 2, \quad (\text{S5})$$

as observed experimentally. In the main text, we see that locally stretching one well to return the initial state to global equilibrium, gives a protocol that does reach the expected thermodynamic bounds.

#### IV. DETAILED DEFINITION OF PROTOCOLS 1 AND 2

A protocol is specified by giving a precise definition of the potential  $U(x, t)$  in Eq. 15 throughout the protocol, for times  $0 \leq t \leq \tau$ . In the parametrization of Eq. 15, we need to specify the functions  $f(t)$  (tilt),  $g(t)$  (barrier height), and  $r(t)$  (stretching of the right well). Below, we give the explicit functions for both protocols.

##### A. Protocol 1

Protocol 1 is defined by specifying the control functions for tilt,  $f_1(t)$ , barrier height,  $g_1(t)$ , and stretching,  $r_1(t)$ . The coordinate is maximally stretched for  $r = 1$ , where it reaches its full amplitude  $\eta$ , while no coordinate stretching is present for  $r = 0$ . These three functions are

$$f_1(t) = \begin{cases} (t/\tau - 0.5)/0.25 & t/\tau \in [0.5, 0.75] \\ 1 & t/\tau \in [0.75, 0.85] \\ 1 - (t/\tau - 0.85)/0.15 & t/\tau \in [0.85, 1] \\ 0 & \text{otherwise} \end{cases} \quad (\text{S6a})$$

$$g_1(t) = \begin{cases} [(t/\tau - 0.5)/0.25]^2 & t/\tau \in [0.25, 0.5] \\ 0 & t/\tau \in [0.5, 0.75] \\ [(t/\tau - 0.75)/0.25]^2 & t/\tau \in [0.75, 1] \\ 1 & \text{otherwise} \end{cases} \quad (\text{S6b})$$

$$r(t) = \begin{cases} (t/\tau - 0.25)/0.25 & t/\tau \in [0, 0.25] \\ 1 & t/\tau \in [0.25, 5] \\ 1 - (t/\tau - 0.5)/0.25 & t/\tau \in [0.5, 0.75] \\ 0 & \text{otherwise} \end{cases}. \quad (\text{S6c})$$

#### B. Protocol 2

Protocol 2 does not involve stretching, meaning that  $r_2(t) = 1$  for all time. The control functions for tilt and barrier height are given by

$$f_2(t) = \begin{cases} (t/\tau - 0.5)/0.25 & t/\tau \in [0.5, 0.75] \\ 1 & t/\tau \in [0.75, 0.85] \\ 1 - (t/\tau - 0.85)/0.15 & t/\tau \in [0.85, 1] \\ 0 & \text{otherwise} \end{cases} \quad (\text{S7a})$$

$$g_2(t) = \begin{cases} [(t/\tau - 0.5)/0.5]^2 & t/\tau \in [0, 1] \\ 1 & \text{otherwise} \end{cases}. \quad (\text{S7b})$$

#### V. CONDITIONED FLUCTUATION RELATION AND ASYMPTOTIC WORK FROM FINITE-TIME MEASUREMENTS IN PROTOCOL 2: THEORY

Starting from the seminal work of Jarzynski [4], one theme of stochastic thermodynamics is that it is possible to estimate equilibrium thermodynamic quantities such as equilibrium free-energy differences from measurements that are conducted on systems out of thermodynamic equilibrium. In this section, we will consider analogous ways to estimate differences in *nonequilibrium* free energies for states that are in local equilibrium, as defined in Eq. 8.

In Protocol 2, we consider a finite time protocol that starts from a full bit and erases it partially using a tilt in

the trajectory. For fixed tilt magnitude, the probability  $p_\tau$  to be in the left well at time  $\tau$  depends on  $\tau$  in a way not known analytically a priori. Moreover, empirical extrapolation to the asymptotic occupation probability,  $\lim_{\tau \rightarrow \infty} p_\tau$ , is not accurate enough, as the uncertainties in estimated probabilities increase with the cycle time  $\tau$ . (Longer cycle times leads to fewer repetitions.) Thus, we cannot extrapolate to infinite  $\tau$ , as we did in Protocol 1.

To estimate asymptotic work using Protocol 2 therefore requires a different strategy that works with measurements performed at a single (large) value of  $\tau$ . Here, we show that by measuring first the average work to carry out the finite-time protocol and then the average work to carry out a time-reversed version of the same protocol, we can deduce the asymptotic minimal average work,  $W = W_\infty$ , i.e.,  $\Delta F_{\text{neq}}$  in Eq. 4 of the main text.

### A. Conditioned fluctuation relation

We start by defining some notation. Let  $P_F(x_0, x_\tau, w)$  be the joint probability for a realization of the forward experiment where a particle starts at position  $x_0$  at time  $t = 0$  and finishes at position  $x_\tau$  at time  $t = \tau$  and for which the changing potential exerts a stochastic work  $w$  on the particle. To be more explicit, consider a path integral over all trajectories  $[x]_0^\tau$  from 0 to  $\tau$  with fixed endpoints and work,

$$P_F(x_0, x_\tau, w) = \int Dx P_F([x]_0^\tau) \times [\delta(x(0) - x_0) \delta(x(\tau) - x_\tau) \delta(w([x]_0^\tau) - w)], \quad (\text{S8})$$

where  $P_F([x]_0^\tau)$  is the probability of the path  $[x]_0^\tau$  and where  $w([x]_0^\tau)$  is the work associated with a given trajectory  $[x]_0^\tau$  (Eq. 16).

Next, assume that the system starts in global equilibrium, which, in the context of this experiment, implies that the system is in local equilibrium within each well, with equal probabilities to be in the two macrostates. The density corresponds to  $\rho_{\text{leq}}(x, t)$  in Eq. 8, with  $p(t) = \frac{1}{2}$ , which we denote  $\rho_{\text{eq}}(x)$ .

For the backward protocol defined by  $U_B(x, t) = U(x, \tau - t)$ , we analogously define the probability  $P_B(x_\tau, x_0, -w)$  that the particle starts at time  $t = 0$  at position  $x_\tau$ , ends at time  $t = \tau$  at position  $x_0$ , and the potential exerts a work  $-w$  on the particle. We further assume that the initial density of the backward protocol is chosen to be the global equilibrium. In the experimental Protocol 2 described in the Methods, the fact that the backward part of the protocol immediately follows the forward part means that the initial density is actually different—local equilibrium with probability  $p_\tau$  for the left well—but this difference turns out not to affect the conditional fluctuation relations that we derive below.

We can relate  $P_F(x_0, x_\tau, w)$  to  $P_B(x_\tau, x_0, -w)$  using the Detailed Fluctuation Relation [5, 6], which generalizes the detailed-balance condition of equilibrium to

nonequilibrium situations. Because the protocol is cyclic, with  $U(x, 0) = U(x, \tau)$ , the usual equilibrium free-energy difference vanishes, and the relation takes a simple form:

$$P_B(x_\tau, x_0, -w) = e^{-\beta w} P_F(x_0, x_\tau, w). \quad (\text{S9})$$

We next marginalize over initial and final positions by integrating  $x_0$  over  $(-\infty, \infty)$  and  $x_\tau$  over  $(-\infty, 0)$ , thereby isolating forward trajectories that end in the left well (state  $L$ ). Imposing in Eq. S9 initial conditioning for the left-hand side and final conditioning for the right-hand side, we then obtain,

$$\begin{aligned} & \underbrace{\left( \int_{-\infty}^0 dx \rho_{\text{eq}}(x) \right)}_{1/2} P_{B|L}(-w) \\ &= e^{-\beta w} \underbrace{\left( \int_{-\infty}^0 dx \rho_{\text{leq}}(x) \right)}_{p_\tau} P_{F|L}(w), \end{aligned} \quad (\text{S10})$$

where  $P_{F|L}(w)$  is the conditional probability to start in global equilibrium, *finish* in  $L$ , and exert work  $w$  in the forward protocol and where  $P_{B|L}(-w)$  is the conditional probability to *start* in  $L$  and exert work  $-w$  in the backward protocol.

Alternatively, we could have integrated over the  $R$  state ( $x > 0$ ) at  $t = \tau$ , to find

$$\begin{aligned} & \underbrace{\left( \int_0^\infty dx \rho_{\text{eq}}(x) \right)}_{1/2} P_{B|R}(-w) \\ &= e^{-\beta w} \underbrace{\left( \int_0^\infty dx \rho_{\text{leq}}(x) \right)}_{1-p_\tau} P_{F|R}(w), \end{aligned} \quad (\text{S11})$$

where, in our slow protocol, the forward process finishes at time  $\tau$  in local equilibrium  $\rho_{\text{leq}}$ .

Thus, we have the “conditioned Crooks” relations,

$$P_{B|L}(-w) = 2p_\tau e^{-\beta w} P_{F|L}(w) \quad (\text{S12a})$$

$$P_{B|R}(-w) = 2(1-p_\tau) e^{-\beta w} P_{F|R}(w). \quad (\text{S12b})$$

A crucial point for us is that, because of the initial conditioning,  $P_{B|L}(-w)$  is identical regardless of whether the backward protocol starts in global equilibrium or in a local equilibrium with arbitrary  $p_\tau$ . Thus, even though we suppose in the Detailed Fluctuation Relation, Eq. S9, that the backward system starts in global equilibrium, Eq. S12a is valid for the backward protocol used in our Protocol 2. Similar statements apply to  $P_{B|R}(-w)$  and Eq. S12b.

If we integrate the conditional Crooks relations in Eq. S12a and b with respect to work, we find the condi-

tional Jarzynski relations,

$$\left\langle e^{-w/k_B T} \right\rangle_{F|L} = \frac{1}{2p_\tau} \quad (S13a)$$

$$\left\langle e^{-w/k_B T} \right\rangle_{F|R} = \frac{1}{2(1-p_\tau)}. \quad (S13b)$$

By Jensen's inequality, we then obtain refinements of the second law in two "conditioned" forms,

$$W_{F|L} \geq k_B T \ln(2p_\tau) \quad (S14a)$$

$$W_{F|R} \geq k_B T \ln(2(1-p_\tau)). \quad (S14b)$$

Analogous relations were experimentally verified in the context of breaking or restoring ergodicity in [7].

Our protocol can be interpreted in the framework of thermodynamics of symmetry breaking and symmetry restoration of Ref.[7]. The first part of the protocol can be interpreted as a symmetry restoration (see the conditioned second principle given in Eq. S.15 of [7], with  $p_i \rightarrow 1/2$ ) followed by a symmetry breaking (Eq. S.7 of [7] with  $p_i \rightarrow p_\tau$ ). By summing these two relations, the equilibrium free energy difference disappears, and we find Eq. S14a (and similarly for Eq. S14b). In Section VI B, we will see that we can also interpret the results of our experiments as testing the conditioned Crooks relation, Eq. S12, directly.

## B. Conditionally Gaussian work fluctuations

We are now interested in the case where the forward and backward conditional probability density functions are Gaussian distributions, as is the case experimentally for slow protocols. Note that in this case, the unconditioned work is not Gaussian but is rather the weighted sum of Gaussians. This fact traces back to the observation that the protocol time  $\tau$  is long enough to reach local equilibrium but short with respect to global equilibrium, which would be achieved by hops over the barrier. Under these assumptions, we can write Eq. S12a as

$$\frac{\exp\left(-\frac{(w+W_{B|L})^2}{2\sigma_{B|L}^2}\right)}{\sqrt{2\pi\sigma_{B|L}^2}} = 2e^{-\beta w} p_\tau \frac{\exp\left(-\frac{(w-W_{F|L})^2}{2\sigma_{F|L}^2}\right)}{\sqrt{2\pi\sigma_{F|L}^2}}. \quad (S15)$$

Rearranging terms gives

$$\begin{aligned} \exp\left(-\frac{(w-W_{F|L})^2}{2\sigma_{F|L}^2} + \frac{(w+W_{B|L})^2}{2\sigma_{B|L}^2} - \beta w + \ln(2p_\tau)\right) \\ = \frac{\sigma_{F|L}}{\sigma_{B|L}}. \end{aligned} \quad (S16)$$

Taking a natural logarithm and isolating terms of the same order in  $w$  gives

$$\begin{aligned} w^2 \left[ -\frac{1}{2\sigma_{F|L}^2} + \frac{1}{2\sigma_{B|L}^2} \right] + w \left[ \frac{W_{F|L}}{\sigma_{F|L}^2} + \frac{W_{B|L}}{\sigma_{B|L}^2} - \beta \right] \\ + \left[ -\frac{W_{F|L}^2}{2\sigma_{F|L}^2} + \frac{W_{B|L}^2}{2\sigma_{B|L}^2} + \ln(2p_\tau) \right] \\ = \ln\left(\frac{\sigma_{F|L}}{\sigma_{B|L}}\right). \end{aligned} \quad (S17)$$

Since Eq. S17 must hold for all values of  $w$ , the prefactors of  $w^2$  and  $w$ , as well as the constant terms, must each vanish separately. For  $w^2$ , we conclude that

$$\sigma_{F|L} = \sigma_{B|L} \equiv \sigma_L. \quad (S18)$$

The  $w$  and constant terms then imply

$$W_{F|L} + W_{B|L} = \beta\sigma_L^2 \quad (S19a)$$

$$W_{F|L}^2 - W_{B|L}^2 = 2\sigma_L^2 \ln(2p_\tau). \quad (S19b)$$

From the ratio of Eq. S19b to Eq. S19a, we deduce that

$$W_{F|L} + W_{B|L} = \beta\sigma_L^2 \quad (S20a)$$

$$W_{F|L} - W_{B|L} = \frac{2}{\beta} \ln(2p_\tau). \quad (S20b)$$

Adding and subtracting Eqs. (S20a) and (S20b) gives

$$W_{F|L} = \frac{\beta}{2}\sigma_L^2 + \frac{1}{\beta} \ln(2p_\tau) \quad (S21a)$$

$$W_{B|L} = \frac{\beta}{2}\sigma_L^2 - \frac{1}{\beta} \ln(2p_\tau). \quad (S21b)$$

A similar argument conditioned on the  $R$  state gives

$$W_{F|R} = \frac{\beta}{2}\sigma_R^2 + \frac{1}{\beta} \ln(2(1-p_\tau)) \quad (S22a)$$

$$W_{B|R} = \frac{\beta}{2}\sigma_R^2 - \frac{1}{\beta} \ln(2(1-p_\tau)). \quad (S22b)$$

Now, by using the law of total probability, we can obtain the average unconditioned forward work  $W_F$  and the average unconditioned backward work  $W_B$ . In Protocol 2, the forward process finishes in local equilibrium. The backward process starts in the same local-equilibrium state, and we obtain

$$W_F = p_\tau W_{F|L} + (1-p_\tau) W_{F|R} \quad (S23a)$$

$$W_B = p_\tau W_{B|L} + (1-p_\tau) W_{B|R}. \quad (S23b)$$

Finally, from Eqs. (S21)–(S23), we have

$$\begin{aligned} \frac{1}{2}(W_F - W_B) &= \frac{1}{2}p_\tau (W_{F|L} - W_{B|L}) \\ &\quad + \frac{1}{2}(1-p_\tau) (W_{F|R} - W_{B|R}) \\ &= \frac{1}{\beta} [p_\tau \ln(2p_\tau) + (1-p_\tau) \ln(2(1-p_\tau))] \\ &= k_B T (\ln 2) [1 - H(p_\tau)], \end{aligned} \quad (S24)$$

which is Eq. 14 in the main text, with  $H(p_\tau)$  in bits. Here, Eq. S24 is valid for states that are in local equilibrium, as defined in Eq. 8.

Thus, by a careful combination of the easily measured average quantities  $W_F$  and  $W_B$  from the forward and backward protocols, the terms involving the variances  $\sigma_L^2$  and  $\sigma_R^2$  cancel, and we can isolate the desired Shannon-entropy term  $H(p_\tau)$ . Intuitively, the average work done is composed of two terms: one from the asymptotic nonequilibrium free energy and one from the fluctuations due to a finite-time protocol. Because of the even character of fluctuations at finite time, half the difference of  $W_F$  and  $W_B$  then corresponds to averaging the nonequilibrium free energy contribution while canceling the dissipation due to a finite-time protocol. We stress that this cancellation works, in general, only for slow protocols where conditional distributions are Gaussian (but no hops over the barrier occur).

## VI. CONDITIONED FLUCTUATION RELATION AND ASYMPTOTIC WORK FROM FINITE-TIME MEASUREMENTS IN PROTOCOL 2: EXPERIMENT

Protocol 2 begins with one bit of information, with the initial probability to be in the left state  $p_0 = \frac{1}{2}$ , and erases a fraction of the information by altering the probability at the end of the protocol to  $p_\tau$ , which is in the range  $0 \leq p_\tau \leq 1$ . Tilting the double-well potential by a small amount would seem a straightforward way to erase a small amount of information. Unfortunately, predicting the tilt needed to bring two states to equilibrium when crossover (mixing) occurs is difficult. Since the crossover time also depends on the rate at which the barrier is lowered, the required tilt varies with cycle time. But varying the tilt also alters  $p_\tau$ , which must then also be extrapolated to long times. These onerous requirements rule out extrapolation as a practical way of measuring the mean work in a protocol with tilt. As a way around this difficulty, we designed Protocol 2 with a small tilt, but we work at a fixed protocol time  $\tau$ . To isolate the minimal average work, we combine both forward and backward manipulations of the potential and use the results derived above in Section V.

Protocol 2 is illustrated in Fig. S3. The first step is to lower the barrier. The next is to tilt the potential by the chosen amplitude,  $A$ . Positive tilt amplitudes ( $A > 0$ ) tend to push the particle to the right and decrease the probability  $p_\tau$  to end up in the *left* well at time  $\tau$ . The last step in the forward part of the protocol is to restore the barrier to the original height of  $10 k_B T$  and untilt.

Figure S4 shows the probability  $p_\tau$  that the system ends up in the left well at time  $\tau$  as a function of tilt amplitude  $A$  (red markers). We note that, for fixed  $A$ , the probability at the end of the protocol depends on its length  $\tau$ . The results presented here and in the main text are all for  $\tau = 2$ . Recall that the cycle time  $\tau$  is scaled by  $\tau_0 = (2x_m)^2/D \approx 10$  s, which is the time for a particle to diffuse the distance between the two local minima, in the absence of a virtual potential. At the end of the

forward part of the protocol, we reverse the changes in the potential. The gray markers in Figure S4 confirm that the protocol is reversible: for all tilt amplitudes  $A$ , the system returns to its initial state with  $p_0 = 0.5$ , and  $H_0 = 1$ . We estimate work using Sekimoto's formula for both the forward and time-reversed protocol sections. (See Methods, Eq. 16.)

In the main text, we assert that we can deduce the form of the Gibbs-Shannon entropy function via measurements of the mean work to carry the forward and backward portions of the protocol (Eq. 5 of the main text), as derived above in Section V B. That derivation assumes, first of all, that conditional work distributions are Gaussian. It also uses intermediate results such as the conditioned Crooks relations, Eq. S12. Here, we give experimental evidence to support these claims and assumptions.

### A. Experimental conditional work distributions are consistent with Gaussian

We begin by showing that the conditional work distributions are consistent with the Gaussian form. Figure S5a shows the measured work distributions for the forward  $P_F(w)$  and backward  $P_B(-w)$  protocols. Although difficult to see explicitly given our resolution, these are *not* expected to be Gaussian but are rather the sum of two conditional Gaussian distributions, depending on whether the particle ends up in the left or right wells (forward protocol) or starts in one of those states (reverse protocol). The solid lines denote the sum of two Gaussian fits to conditional distributions, weighted by the probability for the forward protocol to end up in the left well.

We next use the law of total probability to decompose the work distributions into conditional distributions for a particle ending (or starting) in either well:

$$P_F(w) = p_\tau P_{F|L}(w) + (1 - p_\tau) P_{F|R}(w) \quad (\text{S25a})$$

$$P_B(-w) = p_\tau P_{B|L}(-w) + (1 - p_\tau) P_{B|R}(-w), \quad (\text{S25b})$$

where  $P_{F|L}(w)$ ,  $P_{F|R}(w)$ ,  $P_{B|L}(-w)$ , and  $P_{B|R}(-w)$  are defined in Section V A. Figure S5b and c shows the histogram estimates of the conditional work distributions. The solid lines show that the protocols are slow enough that the empirical conditional work distributions are consistent with Gaussian distributions.

Finally, Figure S6 compares  $W_F$ ,  $W_B$ , and  $H(p_\tau)$ . We stress that the protocols must be executed sufficiently slowly that the conditional work distributions are Gaussian.

### B. Experimental test of conditional Crooks relations

The next step is to show that our experimental results are consistent with the conditional Crooks relations,

Eq. S12. Similar tests have been previously done by Junier et al. [8]. First, we plot the measured conditional work distributions for the backward protocols,  $P_{B|L}(-w)$  and  $P_{B|R}(-w)$ , along with their Gaussian fits. These are the red markers and light red solid line in Fig. S7a and b and reproduce the results of Fig. S5. We then calculate the corresponding forward conditional work distributions  $P_{F|L}(w)$  and  $P_{F|R}(w)$  using Eq. S12 and plot that distribution as the light black curves in Fig. S7a and b. Finally, we plot the measured conditional work distributions and show that they are consistent with values expected. The agreement is tested to a higher precision for a particle ending in the left well (Fig. S7a), because the probability for ending in the left well is  $p_\tau = 0.85 \pm 0.02$ , which implies that there are more work measurements for a particle ending in the left well and, hence, better statistics. In brief, we have shown that the two conditional distributions are related as the conditional Crooks relation asserts they should be.

To test these relations in another way, we sum Eq. S12 and combine with Eq. S25 to find

$$\ln \left( \frac{2P_F(w)}{P_{B|L}(-w) + P_{B|R}(-w)} \right) = \frac{w}{k_B T}. \quad (\text{S26})$$

In Fig. S7c, we plot the left-hand side of Eq. S26 versus work  $w$  (red markers) and confirm the expected linear relation (solid line).

## VII. TESTING THE CONTINUUM VERSION OF THE SHANNON ENTROPY FUNCTION

In the main text, we show experimentally that the system entropy  $S$  that appears in the second law (Eq. 3) is

consistent with the Gibbs-Shannon form of the entropy (Eq. 1). In fact, our experiment also tests a stronger statement: In the Markovian context of Langevin equations such as Eq. 2, the total entropy production  $S_{\text{tot}}$  that appears in Eq. 3 of the main text is equal to [6, 9, 10],

$$S_{\text{tot}} = D_{\text{KL}} \left( P_F ([x]_0^\tau) \mid P_B \left( \overleftarrow{[x]_0^\tau} \right) \right), \quad (\text{S27})$$

where  $P_F ([x]_0^\tau)$  is the probability of the path  $[x]_0^\tau$  under the forward protocol and  $P_B \left( \overleftarrow{[x]_0^\tau} \right)$  is the probability of the same path  $[x]_0^\tau$ , but read backward, for the backward protocol based on the potential  $U^B(x, t) = U(x, \tau - t)$ . The initial density of the backward process is  $\rho(x, \tau)$ .

More precisely, after defining the form of the work  $W$  done on the system to be the average of the Sekimoto formula (Eq. 16), we can use the first law (Eq. 4 of the main text) to deduce the form of the heat  $Q$  exchanged with the medium. The Clausius relation then implies the associated form of the exchanged entropy  $S_m$  that appears in Eq. 3. We can then prove, using Eq. 3 of the main text, that the Gibbs-Shannon form for  $S$  is equivalent to the form of  $S_{\text{tot}}$  given here in Eq. (S27) [6].

Finally, those who consider it “obvious” to use the Gibbs-Shannon form of entropy in the second law (Eq. 3) will perhaps agree that testing the form of  $S_{\text{tot}}$  given in Eq. S27 is less obvious.

- 
- [1] J. Happel and H. Brenner, *Low Reynolds Number Hydrodynamics: With Special Applications to Particulate Media* (Martinus Nijhoff, 1983).
  - [2] Y. Jun, M. Gavrilov, and J. Bechhoefer, “High-precision test of Landauer’s principle in a feedback trap,” *Phys. Rev. Lett.* **113**, 190601 (2014).
  - [3] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, “Dissipation: The phase-space perspective,” *Phys. Rev. Lett.* **98**, 080602 (2007).
  - [4] C. Jarzynski, “Nonequilibrium equality for free energy differences,” *Phys. Rev. Lett.* **78**, 2690–2693 (1997).
  - [5] C. Jarzynski, “Hamiltonian derivation of a detailed fluctuation theorem,” *J. Stat. Phys.* **98**, 77–102 (2000).
  - [6] R. Chétrite and K. Gawędzki, “Fluctuation relations for diffusion processes,” *Commun. Math. Phys.* **282**, 469–518 (2008).
  - [7] É. Roldán, I. A. Martínez, J. M. R. Parrondo, and D. Petrov, “Universal features in the energetics of symmetry breaking,” *Nature Phys.* **10**, 457–461 (2014).
  - [8] I. Junier, A. Mossa, M. Manosas, and F. Ritort, “Recovery of free energy branches in single molecule experiments,” *Phys. Rev. Lett.* **102**, 070602 (2009).
  - [9] C. Maes and K. Netočný, “Time-reversal and entropy,” *J. Stat. Phys.* **110**, 269–310 (2003).
  - [10] C. Maes, “The fluctuation theorem as a Gibbs property,” *J. Stat. Phys.* **95**, 367–392 (1999).

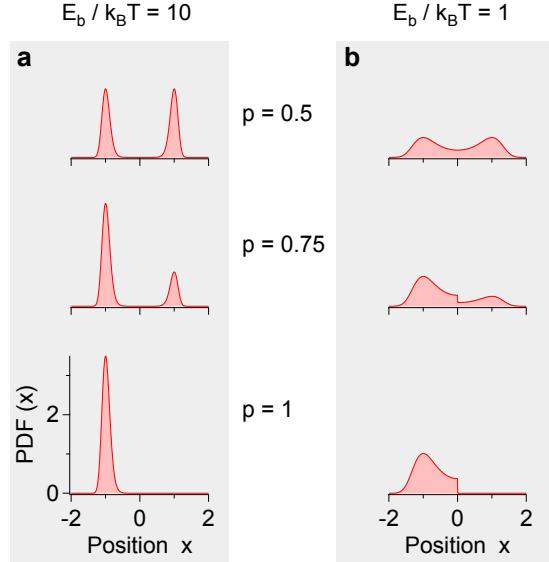


FIG. S1. Effect of dimensionless barrier height  $E_b/k_B T$  on local equilibrium distribution,  $\rho_{\text{eq}}(x)$  for a symmetric, double-well potential. **a**, For a high barrier, there are effectively two separate distributions for all weights  $p(t)$ . **b**, For a low barrier, there is a significant jump discontinuity in density at  $x = 0$  for  $p(t)$  sufficiently different from 0.5.

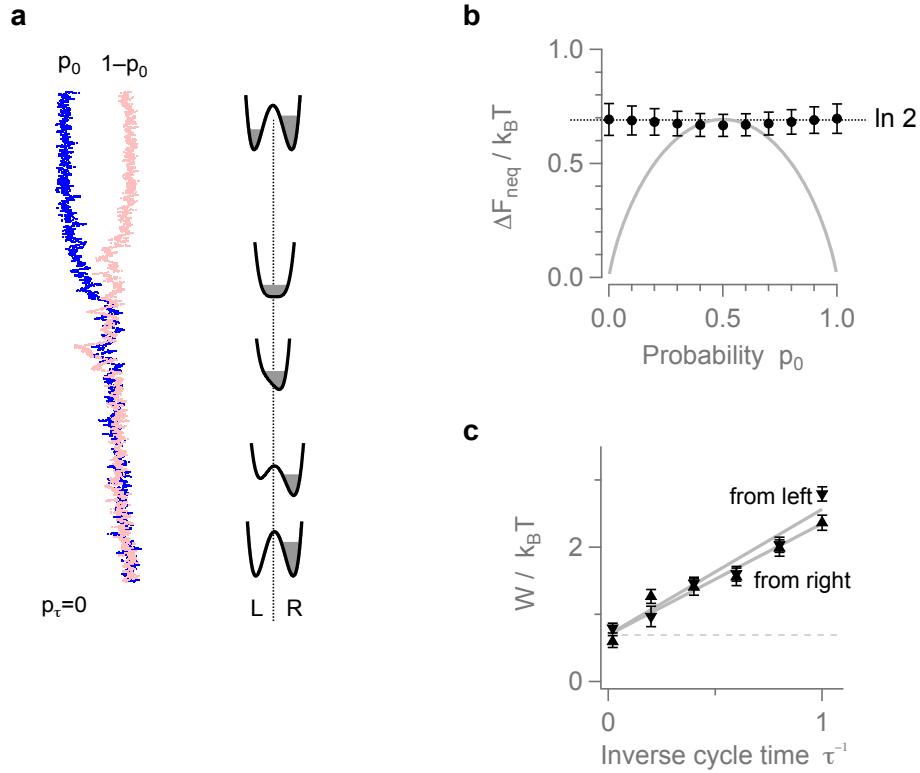


FIG. S2. Naive version of Protocol 1. **a**, Sample trajectories starting from left and right wells, along with potentials at time points within the protocol. **b**, For all initial probabilities  $p_0$ , the average change in nonequilibrium free energy to erase is  $k_B T \ln 2$  for slow protocols. **c**, The average conditional work to erase is asymptotically  $k_B T \ln 2$  for particles that start in the left well and also for particles that start in the right well.

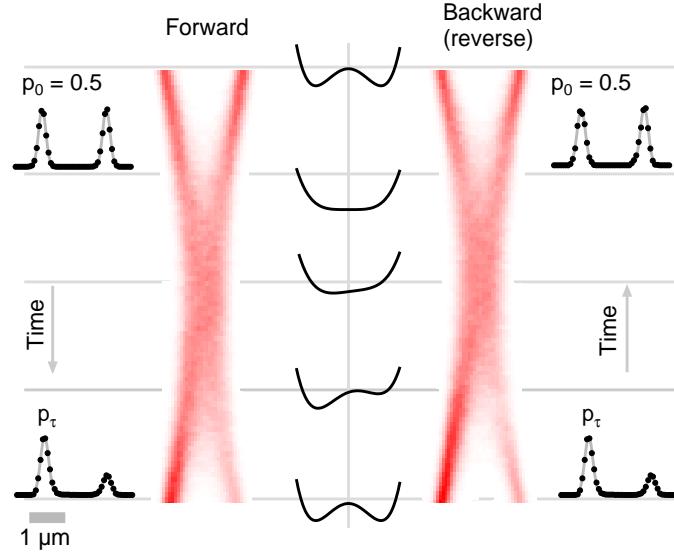


FIG. S3. Protocol 2: Path probability densities for partial erasure in the forward protocol (left) and its accompanying backward protocol (right). In the forward protocol, one bit of information is erased to the left well, with probability  $p_\tau = 0.75 \pm 0.02$  for tilt amplitude  $A = -0.03$ . The duration of the protocol,  $\tau = 2$ , corresponds to a physical time of 20 s. The backward protocol, played forward in time, returns a particle to the initial state with probability 0.5.

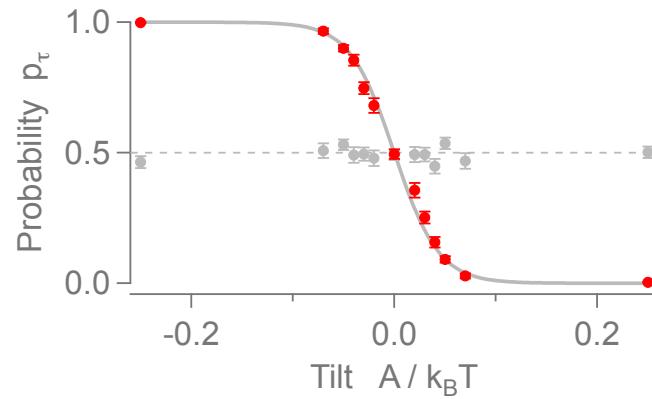


FIG. S4. Erasure probability recorded for different tilt amplitudes. Red markers show probability  $p_\tau$  at the end of the partial erasure experiment, while gray markers show the probability of ending up in the left well for the time-reversed protocol. Solid gray line is empirical function relating the tilt amplitude  $A$  to the probability  $p_\tau$  of being in the  $L$  state at time  $\tau$ , given by  $p_\tau = f(A) = 0.5[1 - \tanh(23A)]$  for  $\tau = 2$ .

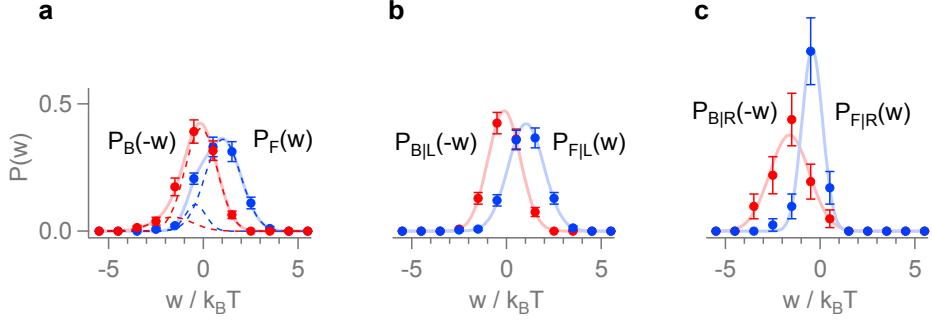


FIG. S5. Conditional work distributions are consistent with Gaussian. **a**, Estimated unconditioned work distributions  $P_F(w)$  and  $P_B(-w)$  for forward and reverse parts of protocols (red and blue markers). Solid lines represent the sum of the corresponding two Gaussian distributions in **b** and **c**. Dashed lines show contributions from weighted conditional Gaussian distributions. **b**, Conditional work distributions for the left state for forward  $P_{F|L}(w)$  and backward  $P_{B|L}(-w)$  protocols. Solid lines are fits to Gaussian distributions. **c**, Same, for right state. Tilt amplitude is set to  $A/k_B T = -0.04$ , and we measure  $p_\tau = 0.85 \pm 0.02$ .

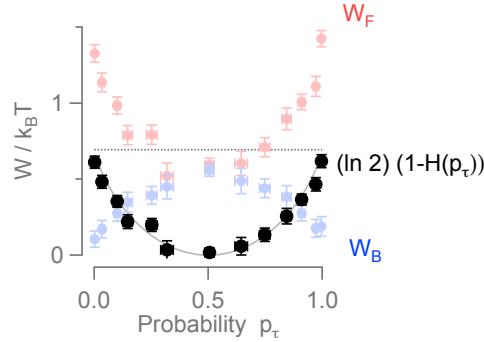


FIG. S6. Mean work in forward protocol  $W_F$  compared with reverse protocol  $W_B$ . The desired Shannon-entropy term  $H(p_\tau)$  is isolated from  $W_F$  and  $W_B$  using Eq. 14 of the main text, or Eq. S24 here.

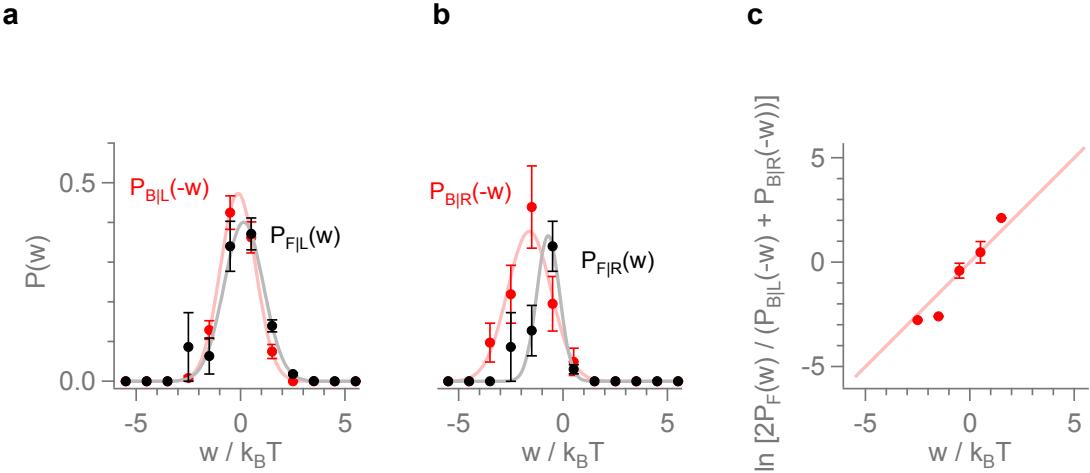


FIG. S7. Comparison between measured and calculated conditional work distributions for Protocol 2. **a**, Conditional work distribution with left-well conditioning. Red markers show measured values, black markers show estimates using Eq. S12a. **b**, Conditional work distribution with right-well conditioning. Red markers show measured values, black markers show estimates using Eq. S12b. **c**, Test of Eq. S26. Solid line has slope = 1.