Editors' Suggestion   Featured in Physics

Bayesian Information Engine that Optimally Exploits Noisy Measurements

Tushar K. Saha[*], Joseph N. E. Lucero[§], Jannik Ehrich, David A. Sivak[‡] and John Bechhoefer[†]
Department of Physics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

We have experimentally realized an information engine consisting of an optically trapped, heavy bead in water. The device raises the trap center after a favorable "up" thermal fluctuation, thereby increasing the bead's average gravitational potential energy. In the presence of measurement noise, poor feedback decisions degrade its performance; below a critical signal-to-noise ratio, the engine shows a phase transition and cannot store any gravitational energy. However, using Bayesian estimates of the bead's position to make feedback decisions can extract gravitational energy at all measurement noise strengths and has maximum performance benefit at the critical signal-to-noise ratio.

Information engines are a new class of engine that use information as fuel to convert heat from a thermal bath into useful energy. They exploit knowledge of thermal fluctuations to apply feedback and extract energy from the thermal bath, while paying costs required by the second law of thermodynamics to process that information [1]. In the past decade, information engines have been realized in a wide range of physical systems [2–8]. For practical application, it is important to understand how to maximize the engine's output [9,10] and efficiency [10–13]. In addition to straightforward measures of performance such as speed of operation or power imparted to a load, there are also more subtle ones such as the ability to maximize information transmission [14,15].

One obstacle that degrades the output of an information engine is inaccurate information about the system that arises from measurement noise. Since information engines respond to measurements of thermal fluctuations, measurement noise can lead to wrong feedback decisions. Feedback actions chosen based on inaccurate measurements reduce the work extracted from the surrounding thermal bath and can even, at high noise levels, lead to a net heating of the thermal bath [16].

Previous efforts to account for noisy measurements in information engines have all used "naive" feedback algorithms based directly on the most recent noisy measurement [16,17]. Here, we show that such information engines, with unidirectional ratchets, have a phase transition between working and nonworking regimes: Below a critical level of signal-to-noise ratio for measurements of the engine state, a "pure" information engine—one that requires no work input beyond that needed to run the measurement and control apparatus—is not possible.

Although previous studies noted the degradation of performance due to measurement noise, they did not attempt to alter the feedback algorithm to compensate. Yet theoretical studies have indicated that incorporating the information contained in past measurements via optimal feedback control could greatly improve the performance of an information engine [17–20]. Indeed, experiments in other areas of physics have used feedback that incorporates Bayesian estimators to demonstrate spectacular results, even in the presence of high measurement noise; significant achievements include trapping a single fluorescent dye molecule that is freely diffusing in water [21] and cooling a nanoparticle to the quantum regime of dynamics [22,23].

In this Letter, we present an experimental realization of an optimal Bayesian information engine that retains the relevant memory of all past measurements in a single summary statistic. Using the extra information from past measurements and correctly compensating for delays in the feedback loop via predictive estimates, we extract and store significant amounts of energy, even in the presence of high measurement noise.

Our implementation of the Bayesian filter uses the optimal affine feedback control algorithm [24], at optimal experimental parameters [10], to maximize the engine's rate of gravitational-energy storage. The relevant information from past observations is used to minimize the uncertainty in the bead's position. This Bayesian information engine extracts energy even at low signal-to-noise ratio (SNR), avoiding the phase transition in the naive information engine that leads to zero output. Under any conditions, this engine extracts at least as much work as the naive engine and reaches the maximal output power possible for Gaussian information engines [25] subject to an affine feedback rule.

Experimental setup.—Our information engine consists of a 4-μm heavy bead trapped in a horizontally aligned optical trap. Because of the surrounding heat bath, the bead's position fluctuates about an equilibrium average. The heavy bead is also subject to gravity. The bead's position is measured with a sampling time $t_s = 20\ \mu$s using the scattered light from a detection laser.
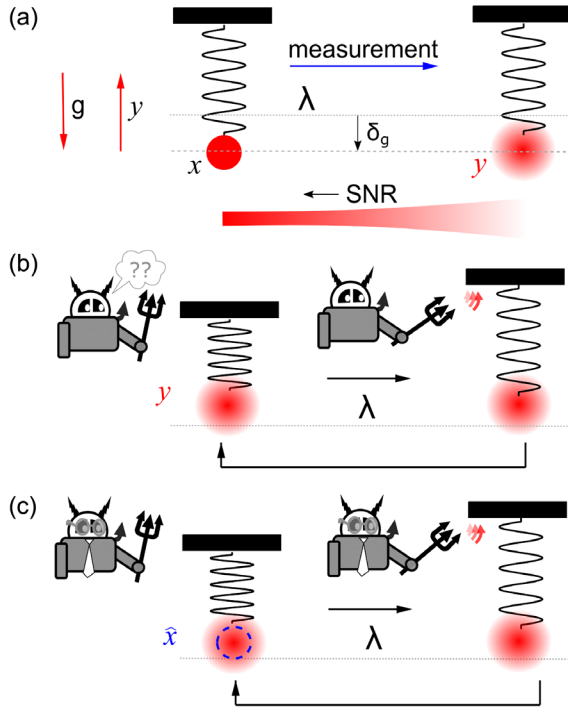
FIG. 1. Schematic information engine. (a) Noisy detector measures position y of the bead, actually located at x. Ratchet based on either (b) noisy position measurement y or (c) Bayesian position estimate $\hat{x}$ (blue dashed circle).

In the operation of the engine, an observed "up" fluctuation increases the bead's gravitational potential energy and can be captured ("rectified") by a quick feedback response that shifts up the trap position. The response takes place after a one-time-step delay $t_d = t_s = 20$ μs, with the shift chosen so that the trap does zero work on the bead. For this "pure" information engine, the work to run the motor is associated only with the measuring device and feedback controller, not the engine itself.

The optically trapped bead can be modeled by a spring-mass system [Fig. 1(a)]. The true position of the bead is estimated from a noisy measurement y. Measurement noise is increased by reducing the intensity of the detection laser beam.

Figure 1(b) shows a "naive" information engine that directly uses a noisy measurement y to apply feedback. By contrast, the "Bayesian" information engine bases its feedback on the best estimate $\hat{x}$ of the bead's position [Fig. 1(c)]. The bead's position is estimated using a Bayesian filter that explicitly models the measurement noise and feedback delay [26].

The information engine can extract energy at high measurement noise because feedback decisions based on filtered estimates of the bead's position are more likely to ratchet to a "true" upward fluctuation rather than to measurement noise.

Equations of motion.—The dynamics of the optically trapped bead obey an overdamped Langevin equation,

$$\gamma \dot{x} = -\kappa[x(t) - \lambda(t)] - mg + \sqrt{2k_B T \gamma} \xi(t), \quad (1)$$

where $x(t)$ is the position at time t of a bead of diameter d and effective (buoyant) mass m in a trap of stiffness κ and center $\lambda(t)$. The friction coefficient is γ, and $\xi(t)$ is a Gaussian random variable with zero mean and covariance $\langle \xi(t)\xi(s) \rangle = \delta(t - s)$. We denote time derivatives by overdots. Scaling all lengths by the equilibrium position standard deviation $\sigma \equiv \sqrt{k_B T/\kappa}$ and times by the bead relaxation time $\tau_r = \gamma/\kappa$ gives the nondimensionalized Langevin equation

$$\dot{x}(t) = -[x(t) - \lambda(t)] - \delta_g + \sqrt{2}\xi(t) \quad (2)$$

for scaled effective mass $\delta_g \equiv mg/(\kappa\sigma)$.

Integrating Eq. (2) between measurements from t to $t + t_s$ gives the discrete dynamics

$$x_{k+1} = x_k e^{-t_s} + (1 - e^{-t_s})(\lambda_k - \delta_g) + \sigma_{t_s}\xi_k \quad (3)$$

for $x_k \equiv x(kt_s)$ and $\lambda_k \equiv \lambda(kt_s)$. The variance $\sigma_{t_s}^2 \equiv 1 - e^{-2t_s}$ of thermal force fluctuations depends on the sampling interval t_s, and ξ_k is a Gaussian random variable with zero mean and covariance $\langle \xi_k \xi_n \rangle = \delta_{kn}$.

The effect of measurement noise is modeled by a measurement variable y_k that is the sum of the bead's true position and additive white Gaussian noise,

$$y_k = x_k + \sigma_m \nu_k, \quad (4)$$

with ν_k a Gaussian random variable with zero mean and covariance $\langle \nu_k \nu_n \rangle = \delta_{kn}$. We also assume that the thermal noise affecting the bead's position is independent of the measurement noise: $\langle \xi_k \nu_n \rangle = 0$, for all k and n.

The trap position λ_k is updated at each time step using a "ratcheting rule,"

$$\lambda_{k+1} = \lambda_k + \Theta(z_k - \lambda_k)[\alpha (z_k - \lambda_k)], \quad (5)$$

where $\Theta(\cdot)$ denotes the Heaviside function, α the scalar feedback gain, and $z_k \in \{y_k, \hat{x}_{k+1}\}$ the estimate of the bead's position (using either the naive measurement y_k or the Bayesian estimate $\hat{x}_{k+1}$). Because of delays, both the naive and Bayesian estimates of position use information from y_k but not y_{k+1}; however, the Bayesian estimate also implicitly incorporates past information $\{y_{k-1}, y_{k-2}, \ldots\}$ and uses the deterministic component of system dynamics to predict y_{k+1} (see Eq. (6) below).

Estimating the bead position.—Since the actual bead position x fluctuates on a scale σ and the measurement noise fluctuates on a scale σ_m, it is convenient to define a signal-to-noise ratio SNR $\equiv \sigma/\sigma_m$ [27]. In the first "naive" approach to designing feedback based on noisy measurements, the feedback rule, Eq. (5), directly uses the measurement y_k to update the trap position λ_{k+1}. Notice

that this method implicitly estimates the position x_{k+1} by y_k. The naive method performs well at high SNR ($\gg 1$) but poorly at low SNR ($\ll 1$), where a unidirectional ratchet (implemented via the Heaviside function) often responds to noise rather than actual bead movements.

In the second "filtering" approach, we improve the estimate of x_{k+1} by using a Fokker-Planck equation to predict the position probability $p(x_{k+1})$ given $p(x_k)$, which is itself calculated from measurements up to y_{k-1}. One then updates (or "corrects") the prediction for time $k+1$ by incorporating the measurement y_k, using Bayes' rule [[28] Sec. III]. For systems evolving according to linear dynamics and subject to Gaussian noise, $p(x_k)$ remains Gaussian for all k (if $p(x_0)$ is initialized as Gaussian, see Supplemental Material [[28] Sec. III]) and can be summarized by update equations for the mean $\hat{x}_k$ and variance. The Bayesian filter is then known as the "predictive Kalman filter," and one finds [compare Eq. (3)] [34]

$$\hat{x}_{k+1} = \underbrace{\hat{x}_k e^{-t_s} + (1 - e^{-t_s})(\lambda_k - \delta_g)}_{\text{predict}} + \underbrace{L\,(y_k - \hat{x}_k)}_{\text{correct}}, \qquad (6)$$

where the scalar filter gain L "corrects" the naive prediction using the difference between the previous estimate and actual observation [[29] Ch. 8]. The gain L is chosen to minimize the variance $\langle(x_k - \hat{x}_k)^2\rangle$ between the true position and its estimate, and the resulting value is a function of the SNR (see Supplemental Material [[28] Sec. IV, Eqs. (S13) and (S14)]). The variance $\langle(x_k - \hat{x}_k)^2\rangle$ is always less than that of the naive estimator [35] (see Supplemental Material [[28] Sec. V]); it is optimal in that it incorporates all relevant past information contained in the (long) time series $\{y_k, y_{k-1}, \ldots\}$, and no other unbiased estimator—linear or not—has lower variance [36].

Engine thermodynamics.—Given an estimate of the bead's position, we infer the thermodynamic quantities that characterize this engine's performance. The rate at which we extract gravitational energy (i.e., change in bead free energy) during the time interval $[t_k, t_{k+1})$ is [24]

$$\Delta F_{k+1} = \delta_g(\lambda_{k+1} - \lambda_k), \qquad (7)$$

and the time-averaged rate of free-energy change (more informally, the "output power") is $\dot{F} = \tau^{-1} \sum_k \Delta F_k$, where $\tau = Nt_s$ is the total duration of an N-step protocol. We estimate the incremental input work (of the trap on the bead) as

$$\Delta W_{k+1} = \frac{1}{2}[(y_{k+1} - \lambda_{k+1})^2 - (y_{k+1} - \lambda_k)^2], \qquad (8)$$

which estimates input work based on the noisy measurement y_k and not on the true position x_k. The Supplemental Material [[28] Sec. VIII] shows that this input-work estimator is unbiased—the work estimator's mean is equal

to the actual average work value—as a result of feedback delay.

Results.—A "pure" information engine has zero input trap power: $\dot{W} = \tau^{-1} \sum_k \Delta W_k = 0$. Let α^* denote the particular value of the feedback gain α for which the trap power is zero. For an "ideal" pure information engine, one with error-free measurements and no feedback delays, the trap work is zero at $\alpha^* = 2$ [10,28]. Physically, this corresponds to translating the trap to a position opposite its minimum, so that the trap energy is unchanged. Feedback delay and measurement noise reduce α^*. For our experimental conditions with delay of one time step and SNR $= 11$, $\dot{W} = 0$ is satisfied at the lower feedback gain $\alpha^* \approx 1.5$ [Figs. 2(a) and 2(b)].

The Bayesian information engine applies feedback based on the filtered predictive estimate of the bead's position. As such, it accounts in its internal model for feedback delays and measurement noise.

Figure 2(a) shows the input trap power, at fixed SNR ($= 11$), as a function of feedback gain α. Despite the delay and finite SNR, the input trap power is zero for feedback gain $\alpha^* = 2$, similar to that of the "ideal" pure information engine. Figure 2(c) illustrates the operation of the Bayesian
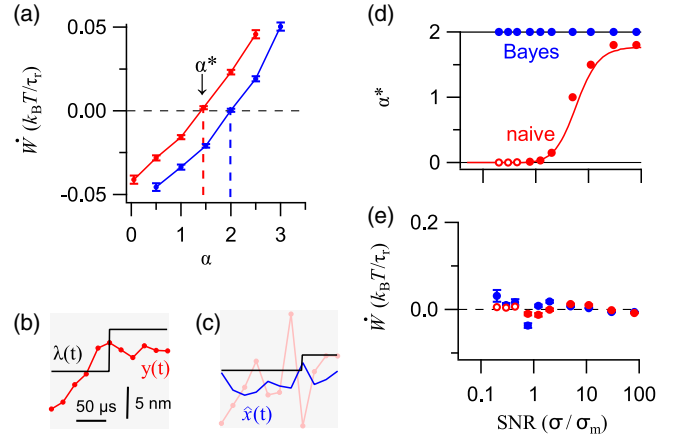

FIG. 2. Tuning the feedback gain α to set trap power $\dot{W} = 0$. (a) Trap power for naive (red) and Bayesian (blue) information engines at fixed SNR $= 11$. (b) Measured bead ($y(t)$, red) and trap ($\lambda(t)$, black) trajectories for the naive information engine at SNR $= 11$. (c) Measured bead ($y(t)$, light red), filtered bead estimate ($\hat{x}(t)$, blue), and trap ($\lambda(t)$, black) trajectories for the Bayesian information engine at SNR $= 2$. (b) and (c) have equal scale bars and satisfy $\dot{W} = 0$. (d) Critical feedback gain α^* and (e) corresponding input trap power, for naive (red) and Bayesian (blue) information engines. Hollow red markers denote SNRs for which $\alpha^* > 0$ could not be found using the procedure outlined in (a). Solid red curve in (d) is from numerical simulation [[28] Sec. VII]. Experiments here and in other figures have sampling frequency $\tau_r/t_s = 41$, trap stiffness $\kappa = 42$ pN/μm, scaled effective mass $\delta_g = 0.8$, diffusion constant $D = 0.12\ \mu$m^2/s, relaxation time $\tau_r = 0.8$ ms, and bead diameter 4 μm. Markers denote experimental means, and error bars the standard errors of the mean.

filter with a trajectory of the Bayesian information engine at lower SNR (= 2). In contrast to the naive information engine, the trap ratchets only when the estimated position (blue) crosses the trap center (black), and not necessarily when the noisy measurement (light red) crosses the trap center.

Next, we investigate how α^* depends on SNR for the naive and Bayesian information engines. For the naive engine, α^* decreases drastically at low SNR [Fig. 2(d)]. For SNR $\lesssim$ SNR$_c = 0.7 \pm 0.1$, where SNR$_c$ denotes the critical vale of the SNR, a nonzero α^* could not be found experimentally using the procedure outlined in Fig. 2(a). The Supplemental Material [[28] Sec. XI] shows that the vanishing of α^* corresponds to a kind of phase transition between a regime where one can set $\dot{F} > 0$ while maintaining $\dot{W} = 0$ and a regime where one cannot.

By contrast, the critical feedback gain α^* remains near 2 for the Bayesian engine [Fig. 2(d)]. The corresponding measured input trap powers for both the naive and Bayesian information engines are close to zero relative to the maximum output power ($\dot{F}_{\max} \approx 0.27 k_B T / \tau_r$) of the engine, at all SNR [Fig. 2(e)].

Taking advantage of predictions in our estimation algorithm thus simplifies the experiments, as it eliminates the need to empirically tune the feedback gain, ensuring that the zero-work condition is always satisfied at $\alpha = 2$. Above, we saw that it also simplifies the work calculations needed to realize a pure information engine, as the value calculated directly from the noisy measurement is an unbiased estimator of the true work.

Finally, Fig. 3(a) compares the performance of the naive and Bayesian information engines, as quantified by the rate of stored gravitational power $\dot{F}$ while keeping $\dot{W} = 0$. Both output powers $\dot{F}$ increase monotonically with SNR and saturate at the same power level at high SNR (> 10).

Although the performance of Bayesian and naive information engines is similar at low and high SNR, there is a striking contrast at intermediate SNR $\lesssim 1$. Indeed, the difference of output powers (Bayesian − naive), normalized by $\dot{F}_{\max}$, significantly exceeds zero for $0.1 \leq$ SNR ≤ 2 and reaches a maximum at SNR $\approx$ SNR$_c$ [Fig. 3(b)].

At high SNR, the Bayesian filter "trusts the observation" and returns an estimate close to the instantaneous measurement, corrected for the expected bias due to the time delay. Since this bias is small for frequent measurements, both engines have similar performance and extract all the favorable thermal fluctuations, saturating at the maximum output power $\dot{F}_{\max} \approx 0.27$. At low SNR, the measurements are so noisy that they exceed the scale of the trap. The Bayesian information engine then extracts negligible power [[28] Sec. IX], while the naive engine extracts zero power. Therefore, at SNR $\gg 1$ and $\ll 1$, the difference of output powers ($\dot{F}_B - \dot{F}_N$) tends to zero. But at intermediate SNR, the effective noise averaging in the Bayesian (Kalman) filter produces more accurate estimates, leading to better feedback decisions and thus improved engine performance.
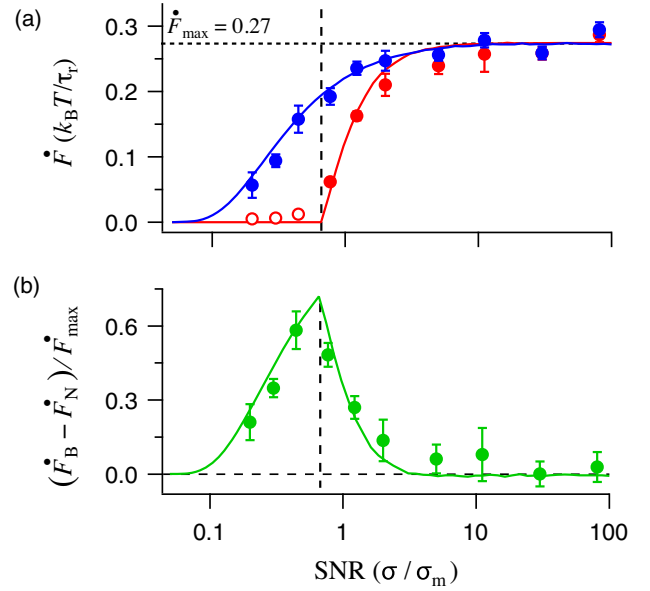

FIG. 3. Performance of the information engines. (a) Output power of naive (red) and Bayesian (blue) information engines as a function of SNR. Hollow red markers denote output power at $\alpha^* = 0$. (b) Difference of output work extraction rates for the Bayesian (B) and naive (N) engines scaled by the maximum rate ($\dot{F}_{\max} = 0.27$). The difference peaks at SNR = SNR$_c \approx 0.7$ (vertical dashed lines). Markers denote experimental means, solid curves the numerical simulations [[28] Sec. VII]. Error bars denote (a) standard error of the mean and (b) propagated standard error of the mean from (a).

To understand why the naive information engine shows a phase transition at a critical signal-to-noise ratio, we numerically solve a self-consistent equation for the SNR at which trap power vanishes. Enforcing the condition that $\alpha^* = 0$ is the unique solution (see Supplemental Material [[28] Sec. X]), we find SNR$_c \approx 0.64$, consistent with both numerical simulations and experiments.

We also find that the phase transition arises from the biased estimate of the bead's position from the noisy measurements. This bias has two origins: the delay due to feedback latency and the failure of the naive measurement to account for the fact that fluctuations above threshold are rare while noise fluctuations of either sign are equally likely. Because fluctuations up to the threshold are rare, the bead is usually below the observed value whenever an apparent threshold crossing is observed.

By contrast, a phase transition does *not* occur for the Bayesian information engine. The Bayesian filter gives an unbiased prediction of the bead's position, accounting for both feedback delay and the "prior" associated with observations near the threshold. As a result, the bead is equally likely to be on either side of the predicted position, allowing one to tune for zero trap power and extract at least some power at *any* SNR value (see Supplemental Material [[28] Sec. XII]).

Conclusion.—Information engines that decide whether to ratchet using single noisy measurements have a phase transition at a critical signal-to-noise ratio SNR_c and cannot function for $\mathrm{SNR} < \mathrm{SNR}_c$. By contrast, if its feedback uses a Bayesian estimate of bead position that incorporates prior measurements, an information engine can operate at *all* values of SNR. The maximum performance benefit over the naive engine occurs at the critical value SNR_c.

The ability to increase the performance of an information engine at low SNR is important for experimental investigations of motor mechanisms that use fluorescent probes [37]. In such applications, lower light intensities for monitoring fluorescent probes reduce photobleaching and allow longer measurements of motor behavior.

In addition, using a filtering algorithm to reduce the required accuracy of information while maintaining a given performance may decrease the thermodynamic costs of processing position measurements. Generally, a lower measurement accuracy reduces the minimum thermodynamic (Landauer) costs of running the controller [16,38]. However, keeping a memory of past observations should increase those costs, as these costs are related to the mutual information the memory stores about the particle position. Storing more measurements results in more information and hence greater costs. In particular, the Bayesian filter uses recursive update relations to implicitly incorporate information from all past measurements. Alternative filters using a finite number of measurements to make an unbiased position estimate would use less information but lead to engines with lower outputs. Further work is needed to compare the efficiency of a feedback strategy that incorporates a memory of past measurements with one based purely on the most recent measurement. Such studies could evaluate the potential performance trade-offs encountered when varying the measurement accuracy.

[*]Corresponding author.
tushars@sfu.ca
[†]Corresponding author.
johnb@sfu.ca
[‡]Corresponding author.
dsivak@sfu.ca
[§]Also at Department of Chemistry, Stanford University, Stanford, California 94305, USA.

[1] J. M. Parrondo, J. M. Horowitz, and T. Sagawa, Thermodynamics of information, Nat. Phys. **11**, 131 (2015).

[2] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano, Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality, Nat. Phys. **6**, 988 (2010).

[3] P. A. Camati, J. P. S. Peterson, T. B. Batalhão, K. Micadei, A. M. Souza, R. S. Sarthour, I. S. Oliveira, and R. M. Serra, Experimental Rectification of Entropy Production by Maxwell's Demon in a Quantum System, Phys. Rev. Lett. **117**, 240502 (2016).

[4] J. V. Koski, A. Kutvonen, I. M. Khaymovich, T. Ala-Nissila, and J. P. Pekola, On-Chip Maxwell's Demon as an Information-Powered Refrigerator, Phys. Rev. Lett. **115**, 260602 (2015).

[5] K. Chida, S. Desai, K. Nishiguchi, and A. Fujiwara, Power generator driven by Maxwell's demon, Nat. Commun. **8**, 15310 (2017).

[6] A. Kumar, T.-Y. Wu, F. Giraldo, and D. S. Weiss, Sorting ultracold atoms in a three-dimensional optical lattice in a realization of Maxwell's demon, Nature (London) **561**, 83 (2018).

[7] J. J. Thorn, E. A. Schoene, T. Li, and D. A. Steck, Experimental Realization of an Optical One-Way Barrier for Neutral Atoms, Phys. Rev. Lett. **100**, 240407 (2008).

[8] M. D. Vidrighin, O. Dahlsten, M. Barbieri, M. S. Kim, V. Vedral, and I. A. Walmsley, Photonic Maxwell's Demon, Phys. Rev. Lett. **116**, 050401 (2016).

[9] T. K. Saha and J. Bechhoefer, Trajectory control using an information engine, in *Optical Trapping and Optical Micromanipulation XVIII* (International Society for Optics and Photonics, Bellingham, WA, 2021), Vol. 11798, p. 117980L, 10.1117/12.2593992.

[10] T. K. Saha, J. N. E. Lucero, J. Ehrich, D. A. Sivak, and J. Bechhoefer, Maximizing power and velocity of an information engine, Proc. Natl. Acad. Sci. U.S.A. **118** (2021).

[11] G. Paneru, D. Y. Lee, T. Tlusty, and H. K. Pak, Lossless Brownian Information Engine, Phys. Rev. Lett. **120**, 020601 (2018).

[12] T. Admon, S. Rahav, and Y. Roichman, Experimental Realization of an Information Machine with Tunable Temporal Correlations, Phys. Rev. Lett. **121**, 180601 (2018).

[13] M. Ribezzi-Crivellari and F. Ritort, Large work extraction and the Landauer limit in a continuous Maxwell demon, Nat. Phys. **15**, 660 (2019).

[14] Y. Tang and A. Hoffmann, Quantifying information of intracellular signaling: Progress with machine learning, Rep. Prog. Phys. **85**, 086602 (2022).

[15] F. Schaper, T. Jetka, and A. Dittrich, Decoding cellular communication: An information theoretic perspective on cytokine and endocrine signaling, Curr. Opin. Endocr. Metab. Res. **24**, 100351 (2022).

[16] G. Paneru, S. Dutta, T. Sagawa, T. Tlusty, and H. K. Pak, Efficiency fluctuations and noise induced refrigerator-to-heater transition in information engines, Nat. Commun. **11**, 1012 (2020).

[17] A. Taghvaei, O. M. Miangolarra, R. Fu, Y. Chen, and T. T. Georgiou, On the relation between information and power in

stochastic thermodynamic engines, IEEE Control Syst. Lett. **6,** 434 (2022).

[18] K. Nakamura and T. J. Kobayashi, Connection Between the Bacterial Chemotactic Network and Optimal Filtering, Phys. Rev. Lett. **126,** 128102 (2021).

[19] J. M. Horowitz, T. Sagawa, and J. M. R. Parrondo, Imitating Chemical Motors with Optimal Information Motors, Phys. Rev. Lett. **111,** 010602 (2013).

[20] N. Rupprecht and D. C. Vural, Predictive Maxwell's demons, Phys. Rev. E **102,** 062145 (2020).

[21] A. P. Fields and A. E. Cohen, Electrokinetic trapping at the one nanometer limit, Proc. Natl. Acad. Sci. U.S.A. **108,** 8937 (2011).

[22] L. Magrini, P. Rosenzweig, C. Bach, A. Deutschmann-Olek, S. G. Hofer, S. Hong, N. Kiesel, A. Kugi, and M. Aspelmeyer, Real-time optimal quantum control of mechanical motion at room temperature, Nature (London) **595,** 373 (2021).

[23] G. P. Conangla, F. Ricci, M. T. Cuairan, A. W. Schell, N. Meyer, and R. Quidant, Optimal Feedback Cooling of a Charged Levitated Nanoparticle with Adaptive Control, Phys. Rev. Lett. **122,** 223602 (2019).

[24] J. N. E. Lucero, J. Ehrich, J. Bechhoefer, and D. A. Sivak, Maximal fluctuation exploitation in Gaussian information engines, Phys. Rev. E **104,** 044122 (2021).

[25] Gaussian information engines are a class of information engines where the system—a particle in our case—is subject to both a linear deterministic force and a Gaussian stochastic force, and where measurements of the system state are contaminated by a different, independent, Gaussian white noise. As such, the Markov dynamical propagator that governs the dynamics of the system probability distribution is also Gaussian.

[26] The delay is also termed "feedback latency," which refers explicitly to the sum of delays in the feedback loop, which include contributions from measurements, computations, and output to the actuator (the acousto-optic deflector, which moves the trap center). We use the more familiar informal language in the text.

[27] SNR is often alternately defined as a ratio of signal-to-noise power, σ^2/σ_m^2.

[28] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.129.130601 for more discussion of the experimental setup, the Bayesian estimator, and the phase transition, which includes Refs. [10,29–33].

[29] J. Bechhoefer, *Control Theory for Physicists* (Cambridge University Press, Cambridge, England, 2021).

[30] A. Kumar and J. Bechhoefer, Nanoscale virtual potentials using optical tweezers, Appl. Phys. Lett. **113,** 183702 (2018).

[31] K. Berg-Sørensen and H. Flyvbjerg, Power spectrum analysis for optical tweezers, Rev. Sci. Instrum. **75,** 594 (2004).

[32] Q. Wang and W. E. Moerner, An adaptive anti-Brownian electrokinetic trap with real-time information on single-molecule diffusivity and mobility, ACS Nano **7,** 5792 (2011).

[33] K. Sekimoto, *Stochastic Energetics* (Springer, New York, 2010).

[34] Note that the predictive Kalman filter first *predicts* the position of the bead using the same model of deterministic dynamics that is present in the particle's equation of motion. It then *corrects* this prediction using the (noisy) observation.

[35] The Bayesian estimator has lower variance than that of the naive estimator provided that the Kalman gain L has the optimal value given by Eq. (S14). If it had a sufficiently different value, the Bayesian estimator variance could exceed that of the naive estimator.

[36] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation* (Prentice Hall, Englewood Cliffs, NJ, 2000).

[37] C. Veigel and C. F. Schmidt, Moving into the cell: Single-molecule studies of molecular motors in complex environments, Nat. Rev. Mol. Cell Biol. **12,** 163 (2011).

[38] J. M. Horowitz and H. Sandberg, Second-law-like inequalities with information and their interpretations, New J. Phys. **16,** 125007 (2014).

Supplementary Material:
Bayesian information engine that optimally exploits noisy measurements

Tushar K. Saha, Joseph N. E. Lucero,* Jannik Ehrich, David A. Sivak, and John Bechhoefer

Department of Physics, Simon Fraser University, Burnaby, BC, V5A 1S6 Canada

CONTENTS

I. EXPERIMENTAL APPARATUS

The experimental apparatus is designed to optically trap a bead in water. The optical trap is built by focusing a green laser using a water-immersion objective (Olympus, 60x 1.2 NA). The position of the trap is controlled using an acousto-optic deflector (AOD, AA opto-electronics). The position of the trapped bead is measured by a red detection laser, propagating anti-parallel to the trapping laser, that is loosely focused in the trapping plane using an air objective. The intensity of the

___

* Department of Chemistry, Stanford University, Stanford, CA 94305, USA

detection bead is controlled using an acousto-optic modulator (AOM, AA opto-electronics) to vary the measurement noise. More details about the experimental apparatus can be found in [1, 2].
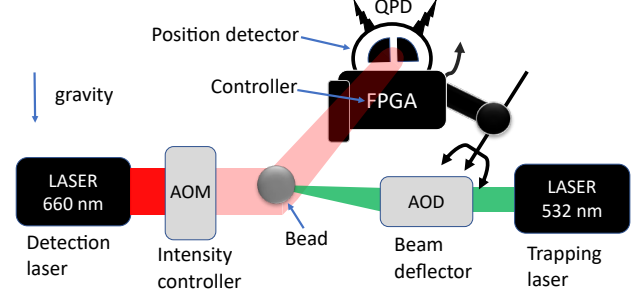

FIG. S1. Schematic diagram of the experimental apparatus.

II. EXPERIMENTAL PARAMETER ESTIMATION

The trap stiffness, diffusion constant, and measurement noise are measured by fitting the power spectrum of the bead's position to a discrete aliased Lorentzian [3]. The discrete equation of motion of the bead can be rewritten as (cf. Eq. (3) in the main text)

$$x_{k+1} = c\,x_k + \Delta x\,\xi_k\,, \tag{S1}$$

where x is measured with respect to $\lambda - \delta_{\mathrm{g}}$. The parameters are

$$c \equiv \exp(-t_{\mathrm{s}}/\tau_{\mathrm{r}}) = \exp(-2\pi f_{\mathrm{c}} t_{\mathrm{s}})\,, \tag{S2}$$

and

$$\Delta x \equiv \left[\frac{(1-c^2)D}{2\pi f_{\mathrm{c}}}\right]^{1/2}\,, \tag{S3}$$

where D is the diffusion constant of the bead, f_{c} the corner frequency, and t_{s} the sampling time. Applying the discrete Fourier transform to the discrete equation of motion of the bead and calculating the expected value of the power spectrum P_f, the discrete aliased Lorentzian is

$$P_f^x = \frac{2(\Delta x)^2 t_{\mathrm{s}}}{1+c^2 - 2c\,\cos 2\pi f t_{\mathrm{s}}}\,, \tag{S4}$$

where f indexes the discrete non-negative frequencies. For noisy measurements

$$y = x + \nu \,, \qquad (S5)$$

where the measurement noise is Gaussian white noise, the power spectrum of y is

$$P_f^y = P_f^x + P^\nu \,, \qquad (S6)$$

where P^ν is the frequency-independent noise floor, and $\sigma_m^2 = P^\nu/2t_s$ the variance of the measurement noise. Equation (S6) uses the property that the measurement noise is independent of both the bead's position x and thermal noise ξ.
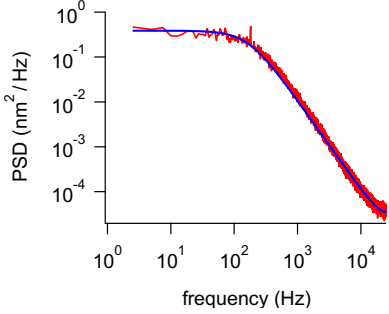

FIG. S2. Power spectral density P_f^y of the measured bead position y in a static trap. Red curve: experiment; blue curve: fit to (S6). The measured fit parameters are $f_c = 167.8 \pm 1.6$ Hz, $D = 0.114 \pm 0.001$ m^2/s, and $\sigma_m = 0.42 \pm 0.06$ nm.

Figure S2 shows the power spectral density of a 4-µm bead in a static trap and the fit to (S6). The trap stiffness is estimated using $\kappa = 2\pi f_c k_B T/D = 40$ pN/µm.

III. PERFORMANCE OF BAYESIAN-FILTER ESTIMATOR

When the measurement noise is additive Gaussian white noise and the system dynamics are linear, as is closely approximated in this experimental setup, an exactly solvable Bayesian filter can be designed to minimize the least-squares error between the filter's state prediction and the true bead position. To derive the optimal filter for the information engine, we note that the bead's motion is a first-order Markov process: its next position x_{k+1} is conditionally independent of its history given its current position x_k and the trap position λ_k, $p(x_{k+1}|x_k, x_{k-1}, \ldots, x_0, \lambda_k) = p(x_{k+1}|x_k, \lambda_k)$. Here, the probability $p(x_{k+1}|x_k, \lambda_k)$ is obtained from Eq. (3),

$$p(x_{k+1}|x_k, \lambda_k) \sim \qquad (S7)$$
$$\mathcal{N}\left(x_{k+1}; x_k \, \mathrm{e}^{-t_s} + \left(1 - \mathrm{e}^{-t_s}\right)(\lambda_k - \delta_g), \, \sigma_{t_s}^2\right) \,,$$

where $\mathcal{N}(x; \mu, \sigma^2)$ denotes a normal distribution over x with mean μ and variance σ^2. The sampling-time dependent variance is $\sigma_{t_s}^2 = 1 - \mathrm{e}^{-2t_s}$. We also assume that the measurements of the bead's position are *memoryless*: the measurement at the current time depends only on the current state, $p(y_k|x_k, x_{k-1}, \ldots, x_0) = p(y_k|x_k)$. Measurements with additive Gaussian white noise satisfy a distribution of the form

$$p(y_k|x_k) \sim \mathcal{N}\left(y_k; x_k, \, \sigma_m^2\right) \,. \qquad (S8)$$

The standard deviation of the noise distribution is the reciprocal of the signal-to-noise ratio, $\sigma_m = 1/\mathrm{SNR}$, and as such can be interpreted as a "noise-to-signal" ratio. Using Bayes' rule, the update distribution is

$$p\left(x_k|Y^k\right) = \frac{p(y_k|x_k)\, p\left(x_k|Y^{k-1}\right)}{p\left(y_k|Y^{k-1}\right)} \,, \qquad (S9)$$

where $Y^k = \{y_k, y_{k-1}, \ldots, y_1\}$ is the set of measurements made up to and including time t_k, and $p\left(y_k|Y^{k-1}\right)$ is a normalizing distribution. From this updated distribution and the dynamics (S7), the predicted distribution at the next discrete time is

$$p\left(x_{k+1}|Y^k\right) = \int \mathrm{d}x_k \; p(x_{k+1}|x_k)\, p\left(x_k|Y^k\right) \,. \qquad (S10)$$

The measurement update and prediction steps for the information engine can be formulated as a sequence of transformations of univariate Gaussian probability densities. Since all densities are Gaussian, we need propagate only their means and variances:

$$\hat{\Sigma}_k^y = \hat{\Sigma}_k + \sigma_{\mathrm{m}}^2 \qquad \text{obs. var.}$$

$$L_{k+1} = \mathrm{e}^{-t_{\mathrm{s}}} \, \frac{\hat{\Sigma}_k}{\hat{\Sigma}_{k+1}^y} \qquad \text{Kalman gain}$$

$$\hat{\Sigma}_k^+ = -\left(L_{k+1}\right)^2 \hat{\Sigma}_{k+1}^y \qquad \text{state var.}$$

$$\hat{x}_k^+ = L_{k+1} \left(y_k - \hat{x}_k\right) \qquad \text{state mean}$$

$$\left.\right\} \text{update} \tag{S11a}$$

$$\hat{\Sigma}_{k+1} = \hat{\Sigma}_k^+ + \mathrm{e}^{-2t_{\mathrm{s}}} \, \hat{\Sigma}_k + \left(1 - \mathrm{e}^{-2t_{\mathrm{s}}}\right) \qquad \text{state var.}$$

$$\hat{x}_{k+1} = \hat{x}_k^+ + \mathrm{e}^{-t_{\mathrm{s}}} \, \hat{x}_k + \left(1 - \mathrm{e}^{-t_{\mathrm{s}}}\right)\left(\lambda_k - \delta_{\mathrm{g}}\right) \qquad \text{state mean}$$

$$\left.\right\} \text{predict} \tag{S11b}$$

Intuitively, the filter generates a prediction mean and variance (S11b), $\hat{x}_{k+1}$ and $\hat{\Sigma}_{k+1}$. The stochastic bead dynamics increase the uncertainty of the bead's position. This position uncertainty is reduced by the information obtained from the previous measurement y_k. The reduction in prediction uncertainty and corresponding change to the prediction mean is achieved by feedback based on the discrepancy between the actual measurement and the prediction, $y_k - \hat{x}_k$. The *Kalman gain L* multiplying this discrepancy accounts for the unreliability of the measurements, given the scale of measurement noise. The information engine uses this filter's prediction mean as the estimate of the bead's position in the feedback decision.

IV.   STEADY-STATE BAYESIAN-FILTER GAIN

In experiments, we assume that the experimental parameters do not change with time. As a result, the means and variances in the time-dependent Eqs. (S11) quickly approach steady-state values. To obtain the steady-state Bayesian-filter gain, we first solve the associated Riccati equation [4, Ch. 8] for the predictive variance,

$$\hat{\Sigma}^{\mathrm{SS}} = \mathrm{e}^{-2t_{\mathrm{s}}} \, \hat{\Sigma}^{\mathrm{SS}} + \left(1 - \mathrm{e}^{-2t_{\mathrm{s}}}\right) - \mathrm{e}^{-2t_{\mathrm{s}}} \, \frac{\left(\hat{\Sigma}^{\mathrm{SS}}\right)^2}{\hat{\Sigma}^{\mathrm{SS}} + \sigma_{\mathrm{m}}^2} \, , \tag{S12}$$

where superscript "SS" without time subscript indicates a steady-state quantity. The relevant solution to this quadratic equation is

$$\hat{\Sigma}^{\mathrm{SS}} = \frac{1}{2}\left(1 - \sigma_{\mathrm{m}}^2\right)\left(1 - \mathrm{e}^{-2t_{\mathrm{s}}}\right) \tag{S13}$$

$$\times \left[1 + \mathrm{sgn}(1 - \sigma_{\mathrm{m}}^2)\sqrt{1 + \frac{4\sigma_{\mathrm{m}}^2}{\left(1 - \sigma_{\mathrm{m}}^2\right)^2 \left(1 - \mathrm{e}^{-2t_{\mathrm{s}}}\right)}}\,\right],$$

where sgn denotes the sign function. We discard the other solution, as it is negative and variances must be positive. The resulting steady-state Bayesian (Kalman) filter gain is

$$L^{\mathrm{SS}} = \frac{\hat{\Sigma}^{\mathrm{SS}}}{\hat{\Sigma}^{\mathrm{SS}} + \sigma_m^2} \, \mathrm{e}^{-t_{\mathrm{s}}} \, . \tag{S14}$$

Finally, the Kalman-filtered position estimate is

$$\hat{x}_{k+1} = L^{\mathrm{SS}}\left(y_k - \hat{x}_k\right) + \mathrm{e}^{-t_{\mathrm{s}}} \, \hat{x}_k + \left(1 - \mathrm{e}^{-t_{\mathrm{s}}}\right)\left(\lambda_k - \delta_{\mathrm{g}}\right) . \tag{S15}$$

V.   VARIANCE OF THE BAYESIAN FILTER

Figure V shows that by choosing the Kalman gain L to have the optimal value (S14), the state variance of the resulting Bayesian estimator (S13) is smaller than that of the naive estimator (see (S16b) below) across all SNR. In fact, no other estimator has a smaller state variance [5].

Because the two noise sources are independent, the state variance of the naive estimator is the sum

$$\hat{\Sigma}_{\mathrm{naive}}^{\mathrm{SS}} = \langle\left(x_{k+1} - y_k\right)^2\rangle \tag{S16a}$$

$$= \sigma_{\mathrm{m}}^2 + \left(1 - \mathrm{e}^{-2t_{\mathrm{s}}}\right) \, , \tag{S16b}$$

where the first term is due to the measurement noise, while the second is the intrinsic variance of an Ornstein-Uhlenbeck process sampled in finite time [6].

In the limit of low SNR ($\sigma_{\mathrm{m}}^2 \to \infty$), the Bayesian estimator's state variance goes to unity, $\hat{\Sigma}^{\mathrm{SS}} \to 1$, while the naive information engine's state variance increases without bound, $\hat{\Sigma}_{\mathrm{naive}}^{\mathrm{SS}} \approx \sigma_{\mathrm{m}}^2 \to \infty$. When the magnitude of the noise is very large, the Bayesian estimator ignores the measurements and relies on the dynamical model (i.e., the prior), resulting in a bounded state variance (σ^2, in physical units). In the limit of high SNR ($\sigma_{\mathrm{m}}^2 \to 0$), the measurements reflect the actual state, and both naive and Bayesian estimators have the same variance, $\hat{\Sigma}^{\mathrm{SS}} \to 1 - \mathrm{e}^{-2t_{\mathrm{s}}}$.
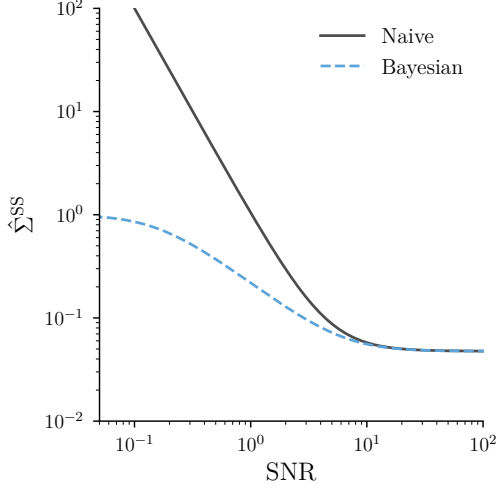
FIG. S3. State variance $\hat{\Sigma}^{\mathrm{SS}} \equiv \left\langle (x_{k+1} - z_k)^2 \right\rangle$ of the naive (solid black curve) and Bayesian (dashed blue curve) estimators as a function of SNR, for sampling frequency $f_\mathrm{s} = 40$.

VI. EXPERIMENTAL IMPLEMENTATION OF THE BAYESIAN FILTER

Experiments on the Bayesian information engine were performed using the Kalman-filtered position estimate obtained from Eq. (S15). Figure S4 shows the performance of the experimental implementation of the Kalman filter for a bead in a static trap. The experiment was performed at SNR ≈ 2, where the experimental parameters were estimated by fitting the power spectral density using Eq. (S6). Figure S4(a) shows the trajectory of the estimated bead position (blue), inferred using Eq. (S15) from the noisy measurements (red).

Figure S4(b) shows the power spectral densities of the noisy measurements (red) and Bayesian estimates (blue). The power spectral density of the noisy measurements saturates at a finite noise floor (PSD = 10^{-3} nm^2/Hz). The filter-estimated position (blue) has a noise floor less then a tenth that of the individual noisy measurements (red). The power spectral density of the innovation $\hat{x}_n - y_n$ is flat, showing that the innovations follow a white-noise distribution, and hence all the position correlations arising from the bead's relaxation in the harmonic trap are extracted by the estimator.

Figure S4(c) shows the normalized autocorrelation of the noisy measurements (red), Bayesian estimates (blue), and innovations (gray). From Eq. (S5), the autocorrelation of the noisy measurements is

$$\langle y_n y_m \rangle = \langle (x_n + \nu_n)(x_m + \nu_m) \rangle \tag{S17a}$$

$$= \sigma_\mathrm{m}^2 \delta_{nm} + \mathrm{e}^{(m-n)t_\mathrm{s}}, \tag{S17b}$$

where we have again used the property that the measurement noise is independent of the true bead position.

The noisy measurements (red curve) show a delta-function peak at $t = 0$ on top of a broader exponential
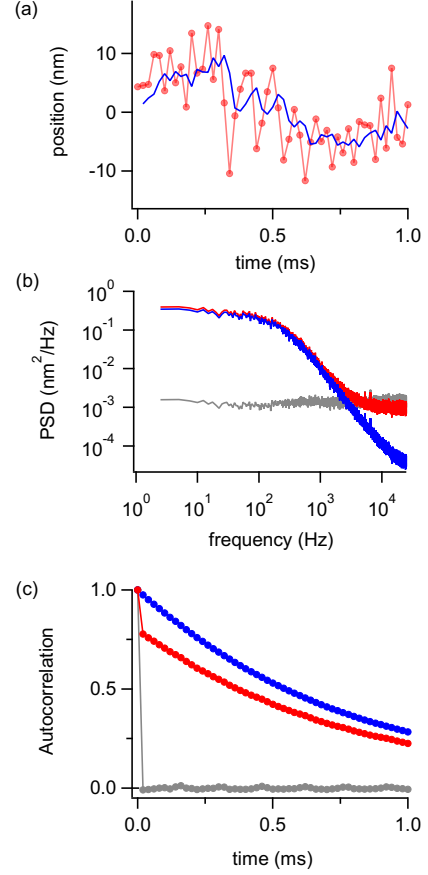

FIG. S4. Experimental performance of the Bayesian-filtered estimate. (a) Noisy experimental measurement (y, red) and the corresponding Bayesian estimate using Eq. (S15) ($\hat{x}$, blue) of the bead's position. (b) Power spectral density and (c) autocorrelation of the noisy measurement (red), Bayesian estimate (blue), and innovations $\hat{x} - y$ (gray). The experiments were performed for SNR ≈ 2, equilibrium standard deviation $\sigma = 10.1$ nm, and measurement noise $\sigma_\mathrm{m} = 4.7$ nm.

decay, reflecting two processes: First, there is the contribution of measurement noise, which is delta correlated; second, there is the contribution of the dynamics, which has an exponential decay of correlations.

By contrast, the Bayesian estimate (blue curve) has only an exponential decay due to the dynamics—as would the autocorrelation function of noise-free dynamics. In other words, the Bayesian estimator distills the dynamical information in the correlated signal. The innovations (gray curve) are almost entirely delta correlated. Indeed, the degree to which the innovations approximate a delta function serves as a measure of the quality of the Bayesian filter. For example, deviations between a parameter value in the physical system and the corresponding model used for the Bayesian filter produce dynamical correlations in the innovations autocorrelation function [4]. Indeed, this mismatch has been used as a way to empirically tune parameters for Bayesian (Kalman)

filters ("innovation whitening") [7].

VII. NUMERICAL-SIMULATION METHODS

The numerical simulations generate multiple long trajectories of the bead using the discrete-time dynamics,

$$x_{k+1} = x_k \, \mathrm{e}^{-t_\mathrm{s}} + \left(1 - \mathrm{e}^{-t_\mathrm{s}}\right) \left(\lambda_k - \delta_\mathrm{g}\right) + \sigma_{t_\mathrm{s}} \xi_k \;, \quad (\text{S18})$$

where $x_k \equiv x(kt_\mathrm{s})$ and $\lambda_k \equiv \lambda(kt_\mathrm{s})$. The corresponding noisy measurement is $y_k = x_k + \sigma_\mathrm{m}\nu_k$, where ν_k is a Gaussian random variable with zero mean and covariance $\langle \nu_k \nu_n \rangle = \delta_{kn}$. For the naive information engine, the trap position is updated using the feedback rule

$$\lambda_{k+1} = \lambda_k + \Theta\left(y_k - \lambda_k\right) \alpha(y_k - \lambda_k). \quad (\text{S19})$$

For the Bayesian information engine, the trap position is updated using the feedback rule

$$\lambda_{k+1} = \lambda_k + \Theta\left(\hat{x}_{k+1} - \lambda_k\right) \alpha(\hat{x}_{k+1} - \lambda_k) \;, \quad (\text{S20})$$

using the Bayesian estimate (S15) of the bead's position. For each engine type, the free-energy change is estimated using Eq. (7) and the work using the unbiased estimator Eq. (8) in the main text.

VIII. TRAP-WORK ESTIMATES ARE UNBIASED

Here, we show that our empirical estimator for the input trap work

$$W_{k+1} = \frac{1}{2}\left[\left(y_{k+1} - \lambda_{k+1}\right)^2 - \left(y_{k+1} - \lambda_k\right)^2\right] \quad (\text{S21})$$

is an unbiased estimator when the feedback rule follows

$$\lambda_{k+1} = \lambda_k + \Theta\left(z_k - \lambda_k\right) \alpha(z_k - \lambda_k) \;. \quad (\text{S22})$$

Note that this feedback rule updates the value of the control parameter $\lambda_k \to \lambda_{k+1}$ at time t_{k+1} according to the estimate z_k of the bead's position at the *previous* time t_k, not the measurement that is made at the current time t_{k+1}. In other words, there is a delay of one cycle between when a measurement is taken and when the corresponding feedback is applied.

For convenience, we introduce a new notation where the superscript on the thermodynamic quantities (e.g., W_k^x) denotes the type of measurement used in the estimation of the thermodynamic quantity. In this notation, we write (S21) as

$$W_{k+1}^y = \frac{1}{2}\left[\left(y_{k+1} - \lambda_{k+1}\right)^2 - \left(y_{k+1} - \lambda_k\right)^2\right] \;, \quad (\text{S23})$$

and the true estimate of the work from stochastic thermodynamics [8] given by the actual potential-energy change when updating λ_k to λ_{k+1},

$$W_{k+1}^x = \frac{1}{2}\left[\left(x_{k+1} - \lambda_{k+1}\right)^2 - \left(x_{k+1} - \lambda_k\right)^2\right] \;. \quad (\text{S24})$$
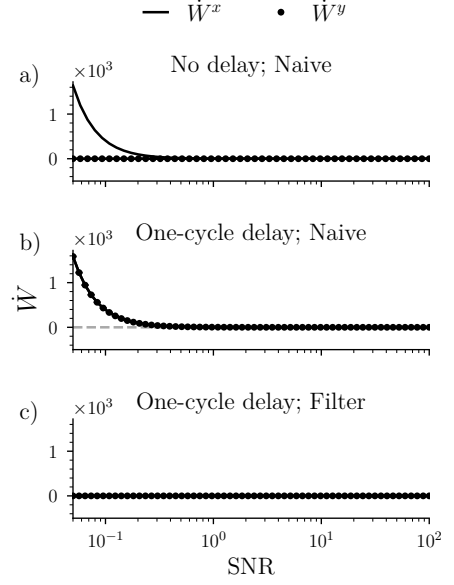

FIG. S5. Numerical simulations of bias in the trap-power estimates from noisy measurements, as a function of SNR. a) Estimated trap power in naive information engine with no feedback delay (latency) and applying feedback based on the most recent noisy measurement: $\lambda_{k+1} = \lambda_k + \Theta\left(y_{k+1} - \lambda_k\right)\left[\alpha(y_{k+1} - \lambda_k)\right]$. b), c) Estimated trap power when feedback is applied after a delay of one time step, using the feedback rule (5) in the main text, based on (b) biased naive $z_k = y_k$ or (c) unbiased filtered $z_k = \hat{x}_{k+1}$ position estimates. Solid curve: estimate using true positions x. Markers: estimate using measurements y_k. The numerical simulations were performed for sampling frequency $f_\mathrm{s} = 40$, feedback gain $\alpha = 2$, and scaled mass $\delta_\mathrm{g} = 0.84$.

Now we compare the empirical estimates to the true values of the thermodynamic quantities. Expanding the definition of the empirical input work (S21) using $y_k = x_k + \nu_k$, where ν_k is a zero-mean Gaussian random variable with variance σ_m^2, gives

$$W_{k+1}^y = \frac{1}{2}\left[\left(y_{k+1} - \lambda_{k+1}\right)^2 - \left(y_{k+1} - \lambda_k\right)^2\right] \quad (\text{S25a})$$

$$= \frac{1}{2}\left[\left(x_{k+1} - \lambda_{k+1} + \nu_{k+1}\right)^2 \right. \quad (\text{S25b})$$

$$\left. - \left(x_{k+1} - \lambda_k + \nu_{k+1}\right)^2\right]$$

$$= \frac{1}{2}\left[\left(x_{k+1} - \lambda_{k+1}\right)^2 - \left(x_{k+1} - \lambda_k\right)^2\right] \quad (\text{S25c})$$

$$+ \left(\lambda_{k+1} - \lambda_k\right)\nu_{k+1}$$

$$= \left(W_{k+1}\right)^x + \left(\lambda_k - \lambda_{k+1}\right)\nu_{k+1} \;. \quad (\text{S25d})$$

Averaging over the noise at time t_{k+1}, when the trap position has a nonnegative update $\lambda_{k+1} - \lambda_k > 0$ (when $\lambda_{k+1} = \lambda_k$, the work is trivially $W = 0$ regardless of

which position estimate is used), we find that

$$\left\langle W_{k+1}^{y}\right\rangle_{\nu} = \left\langle W_{k+1}^{x}\right\rangle_{\nu} + \left\langle (\lambda_k - \lambda_{k+1})\,\nu_{k+1}\right\rangle \quad \text{(S26a)}$$

$$= \left\langle W_{k+1}^{x}\right\rangle_{\nu} + \left\langle [\lambda_k - (\lambda_k + \alpha(z_k - \lambda_k))]\,\nu_{k+1}\right\rangle \quad \text{(S26b)}$$

$$= \left\langle W_{k+1}^{x}\right\rangle_{\nu} - \alpha\left\langle z_k\,\nu_{k+1}\right\rangle + \alpha\left\langle \lambda_k\,\nu_{k+1}\right\rangle \quad \text{(S26c)}$$

$$= \left\langle W_{k+1}^{x}\right\rangle_{\nu}\,, \quad \text{(S26d)}$$

where, in the last step, we have used the property that both the position estimate z_k (which denotes the noisy measurement y_k in the naive case and the predictive estimate $\hat{x}_{k+1}$ in the Bayesian case), and the trap position λ_k at time step k are independent of the *future* measurement noise ν_{k+1}. Without the delay, the variables are correlated and the measurements are biased (Fig. S5a). This calculation suggests that feedback delay actually eliminates the bias in the empirical estimate of the trap work. Using numerical simulations, we compute the empirical power $\dot{W}^{y}$ and find that it agrees well with the true values $\dot{W}^{x}$ for both the naive (Fig. S5b) and Bayesian (Fig. S5c) information engines.

We also note that the trap work could alternatively be estimated using the Bayesian filtered position; however, the estimator in Eq. (S21) is *robustly unbiased*: It is unbiased even if the dynamical model used in the Kalman filter uses parameters that differ from those describing the physical system. The calculation of the trap position λ_k is based on past information and—whatever the algorithm used to compute its value—will always be independent of the current measurement.

IX. EXPERIMENT AND SIMULATION OF THE CRITICAL FEEDBACK GAIN α^{*}

Figure S6 presents Fig. 2 in the main text, but with a logarithmic ordinate. According to simulations, for the naive information engine the critical feedback gain $\alpha^{*} = 0$ is the only solution below a critical signal-to-noise ratio SNR_{c}, as shown in Fig. S6(a). At low SNR, experimental determinations of the critical feedback gain α^{*} do not agree with numerical simulations (Fig. S6(a)), denoted as red hollow markers. Although experiments find $\alpha^{*} > 0$, the naive output powers $\dot{F}$ are still an order of magnitude smaller than those of the Bayesian information engine (Fig. S6(b)). Figure S6(c) shows the corresponding input trap powers $\dot{W}$, which remain small at low SNR.

X. THEORETICAL ARGUMENTS FOR THE PHASE TRANSITION IN THE NAIVE INFORMATION ENGINE

Empirically, the naive estimator can extract free energy only above a critical signal-to-noise ratio $\text{SNR}_{\text{c}} \approx$ 0.7.

We define the critical SNR as the highest value for which $\alpha^{*} = 0$ is the only solution for which $\dot{W} = 0$.
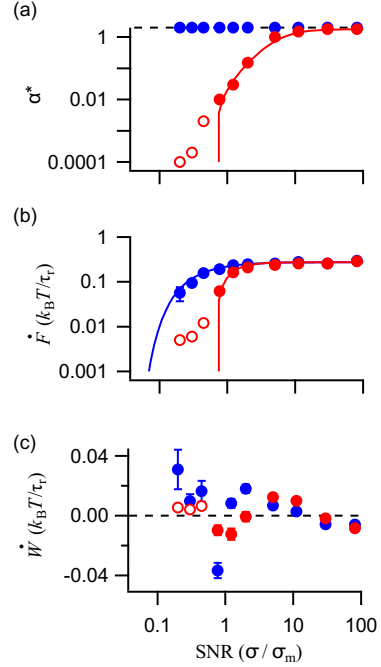

FIG. S6. Experiments (markers) and simulations (curves) from Fig. 2 in the main text, presented on a logarithmic ordinate. (a) The critical feedback gain α^{*}, and the corresponding (b) output power $\dot{F}$ and (c) trap power $\dot{W}$ of the naive (red) and Bayesian (blue) information engines, each as a function of SNR. Hollow markers denote experiments at $\text{SNR} < \text{SNR}_{\text{c}}$. Error bars denote the standard error of the mean.

Note that $\alpha = 0$ corresponds to never moving the trap and therefore not doing any work on the bead. When no feedback is performed, the stationary distribution of the bead inside the trap is the equilibrium Boltzmann distribution,

$$p(x_k|\lambda_k) = \mathcal{N}(x_k;\lambda_k - \delta_{\text{g}},\,1)\,. \quad \text{(S27)}$$

This equation, the bead's propagator $p_x(x_{k+1}|x_k,\lambda_k)$ in Eq. (S7), and the measurement distribution $p_y(y_k|x_k)$ in Eq. (S8) together allow us to evaluate the input work,

$$\langle \Delta W_{k+1} \rangle = \frac{1}{2} \left\langle (x_{k+1} - \lambda_{k+1})^2 - (x_{k+1} - \lambda_k)^2 \right\rangle \tag{S28a}$$

$$= \frac{1}{2} \left\langle [x_{k+1} - \lambda_k - \Theta(y_k - \lambda_k)\,\alpha(y_k - \lambda_k)]^2 - (x_{k+1} - \lambda_k)^2 \right\rangle \tag{S28b}$$

$$= \frac{1}{2} \int d\lambda_k\, p(\lambda_k) \int dx_{k+1} \int dx_k \int_{\lambda_k}^{\infty} dy\; p_x(x_{k+1}|x_k,\lambda_k)\, p_y(y_k|x_k)\, p(x_k|\lambda_k)$$
$$\times \left\{ [x_{k+1} - \lambda_k - \alpha\,(y_k - \lambda_k)]^2 - (x_{k+1} - \lambda_k)^2 \right\} \tag{S28c}$$

$$= \frac{1}{2} \int d\lambda_k\, p(\lambda_k) \int d\tilde{x}^+ \int d\tilde{x} \int_0^{\infty} d\tilde{y}\; p_x(\tilde{x}^+ + \lambda_k|\tilde{x} + \lambda_k, \lambda_k)\, p_y(\tilde{y} + \lambda_k|\tilde{x} + \lambda_k)\, p(\tilde{x} + \lambda_k|\lambda_k)$$
$$\times \left[(\tilde{x}^+ - \alpha\tilde{y})^2 - (\tilde{x}^+)^2 \right] \tag{S28d}$$

$$= \frac{1}{2} \int d\lambda_k\, p(\lambda_k) \int d\tilde{x}^+ \int d\tilde{x} \int_0^{\infty} d\tilde{y}\, \mathcal{N}\left[\tilde{x}^+; \tilde{x}\, e^{-t_s} - \delta_g(1 - e^{-t_s}),\, 1 - e^{-2t_s} \right] \mathcal{N}(\tilde{y}; \tilde{x}, \sigma_m^2)\, \mathcal{N}(\tilde{x}; -\delta_g, 1)$$
$$\times \left[(\tilde{x}^+ - \alpha\tilde{y})^2 - (\tilde{x}^+)^2 \right] \tag{S28e}$$

$$= \frac{\alpha}{4} \left[\alpha\left(\delta_g^2 + 1 + \sigma_m^2\right) - 2\delta_g^2 - 2\,e^{-t_s} \right] \left[1 - \mathrm{erf}\left(\frac{\delta_g}{\sqrt{2(1 + \sigma_m^2)}} \right) \right] + \sqrt{\frac{1 + \sigma_m^2}{8\pi}}\, \alpha(2 - \alpha)\delta_g \exp\left[-\frac{\delta_g^2}{2(1 + \sigma_m^2)} \right], \tag{S28f}$$

where we substituted $\tilde{x} = x_k - \lambda_k$, $\tilde{y} = y_k - \lambda_k$, and $\tilde{x}^+ = x_{k+1} - \lambda_k$ in line (S28d), and used the normalization of the unknown $p(\lambda_k)$ in line (S28f).

Solving $\langle \Delta W_{k+1} \rangle = 0$ for α yields solutions $\alpha^* = 0$, and

$$\alpha^* = 2\, \frac{ \sqrt{2(1 + \sigma_m^2)}\, \delta_g \exp\left[-\frac{\delta_g^2}{2(1 + \sigma_m^2)} \right] - \sqrt{\pi}\left(\delta_g^2 + e^{-t_s} \right)\left[1 - \mathrm{erf}\frac{\delta_g}{\sqrt{2(1 + \sigma_m^2)}} \right] }{ \sqrt{2(1 + \sigma_m^2)}\, \delta_g \exp\left[-\frac{\delta_g^2}{2(1 + \sigma_m^2)} \right] - \sqrt{\pi}\left(\delta_g^2 + 1 + \sigma_m^2 \right)\left[1 - \mathrm{erf}\frac{\delta_g}{\sqrt{2(1 + \sigma_m^2)}} \right] }. \tag{S29}$$

At the critical σ_m^2, the solution α^* bifurcates. We determine the critical SNR from Eq. (S29) by solving

$$\alpha^* = 0 \tag{S30}$$

for σ_m^2, which can only be done numerically. For $t_s = 1/41$ and $\delta_g = 0.8$, the parameters used in Figs. 2 and 3,

$$\mathrm{SNR_c} = (\sigma_m^c)^{-1} \approx 0.64\,. \tag{S31}$$

Figure 2 in the main text shows that this critical SNR indeed empirically coincides with the emergence of a positive α^*. We note that the accuracy of this theoretical estimate (and thus the numerical precision we give) is limited by the accuracy of experimental parameters such as δ_g and t_s which depend on the bead size.

XI. EXPERIMENTAL EVIDENCE SUPPORTING A PHASE TRANSITION IN THE NAIVE INFORMATION ENGINE

In this section, we present detailed experimental evidence for a phase transition in the naive information engine. Figure S7 (a) shows the trap power as a function of the feedback gain α for different SNR, where the markers in the blue shaded region represent SNR values for which only positive trap powers ($\dot{W} > 0$) were measured. The experimental points are fit to a quadratic function; Fig. S7 (b) shows the linear coefficients of the fits in Fig. S7 (a) as a function of SNR. The measured slope switches from negative at high SNR $\gtrsim 0.8$ to positive at SNR $\lesssim 0.6$, consistent with $\mathrm{SNR_c}$ determined in Eq. (S31). To measure the critical signal-to-noise ratio (where the linear coefficient is zero), the linear coefficients for SNRs 0.4–0.8 are linearly fit, giving an estimated zero linear coefficient and hence $\mathrm{SNR_c}$ at 0.7 ± 0.1. In the naive information engine experiments,
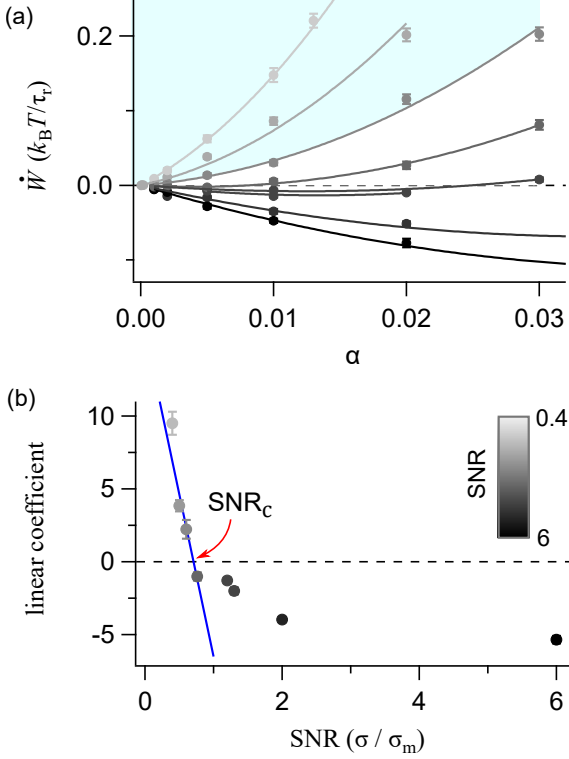
FIG. S7. Phase transition in naive information engine. (a) Trap power as a function of feedback gain α for different signal-to-noise ratios: 0.4, 0.5, 0.6, 0.8, 1.2, 1.3, 2, and 6 (different shades of gray, scale bar in (b)). Markers: experiment; curves: fit to quadratic functions. Blue shaded region: SNR values for which $\dot{W} > 0$ for all values of α. (b) Linear coefficients from quadratic fits (curves in (a)), for different SNRs. Blue line: linear fit to first four points. Error bars each denote standard error of the mean in (a) and parameter fit error in (b).

for SNR < SNR_c, we set $\alpha > \alpha^*$ because we cannot find an α that leads to $\dot{W} < 0$. Thus, the α^* we report in such cases (red hollow markers in Fig. 3, main text) is an upper bound on the α^* that satisfies $\dot{W} \gtrsim 0$: Either there is a smaller positive α^* that we could not detect experimentally because differences in $\dot{W}$ are too close to zero, or the only value that (always) enforces $\dot{W} = 0$ is $\alpha^* = 0$. The corresponding $\dot{F}$ measured using the upper-bound value of α^* then is an upper bound on the output power for SNR < SNR_c.

Importantly, the experimental trap powers in Fig. S7 need not match the work in Eq. (S28f) because that equation assumed that the bead's stationary-state position distribution is the equilibrium distribution (S27), which only strictly holds when $\alpha = 0$.

XII. THE BAYESIAN INFORMATION ENGINE HAS NO PHASE TRANSITION

Here, we present theoretical arguments to support the claim that there is no corresponding "phase transition" for the information engine that estimates the bead position using filtered measurements.

A. Feedback gain

First, we show that feedback gain $\alpha = 2$ generally produces vanishing input work. Using the predictive Kalman filter [Eq. (6)] in the feedback rule [Eq. (5)] with $\alpha = 2$, the trap work is

$$\langle \Delta W_{k+1} \rangle = \frac{1}{2} \left\langle (x_{k+1} - \lambda_{k+1})^2 - (x_{k+1} - \lambda_k)^2 \right\rangle \tag{S32a}$$

$$= \frac{1}{2} \left\langle [x_{k+1} - \lambda_k - 2\,\Theta(\hat{x}_{k+1} - \lambda_k)\,(\hat{x}_{k+1} - \lambda_k)]^2 - (x_{k+1} - \lambda_k)^2 \right\rangle \tag{S32b}$$

$$= \frac{1}{2} \int \mathrm{d}\lambda_k\, p(\lambda_k) \int \mathrm{d}x_{k+1} \int_{\lambda_k}^{\infty} \mathrm{d}\hat{x}_{k+1}\, p(x_{k+1}, \hat{x}_{k+1}|\lambda_k) \left\{ [x_{k+1} - \lambda_k - 2\,(\hat{x}_{k+1} - \lambda_k)]^2 - (x_{k+1} - \lambda_k)^2 \right\} \tag{S32c}$$

$$= 2 \int \mathrm{d}\lambda_k\, p(\lambda_k) \int_{\lambda_k}^{\infty} \mathrm{d}\hat{x}_{k+1}\, p(\hat{x}_{k+1}|\lambda_k)\, (\hat{x}_{k+1} - \lambda_k) \underbrace{\int \mathrm{d}x_{k+1}\, p(x_{k+1}|\hat{x}_{k+1}, \lambda_k)\, (x_{k+1} - \hat{x}_{k+1})}_{=\,0} \tag{S32d}$$

$$= 0\,, \tag{S32e}$$

where (S32d) used the fact that the posterior distribution $p(x_{k+1}|\hat{x}_{k+1}, \lambda_k)$ of the true bead position x_{k+1} is

symmetric around the filter estimate $\hat{x}_{k+1}$, which in this case is the mean of the Gaussian posterior. Figure S8

shows the input work as a function of the feedback gain α for different signal-to-noise ratios. $\alpha = 2$ makes the input work vanish.

B. Free-energy gain

With feedback gain $\alpha = 2$ and the feedback rule in Eq. (5), the free energy-gain per time step, Eq. (7), becomes

$$\langle \Delta F_{k+1} \rangle = \delta_{\mathrm{g}} \langle \lambda_{k+1} - \lambda_k \rangle \tag{S33a}$$

$$= 2\delta_{\mathrm{g}} \int \mathrm{d}\lambda_k \, p(\lambda_k) \underbrace{\int_{\lambda_k}^{\infty} \mathrm{d}\hat{x}_{k+1} \, p(\hat{x}_{k+1}|\lambda_k) \, (\hat{x}_{k+1} - \lambda_k)}_{>\,0} \tag{S33b}$$

$$> 0\,, \tag{S33c}$$

where we used the fact that the probability density $p(\hat{x}_{k+1}|\lambda_k)$ is positive and always has a tail (however small) reaching into the region $\hat{x}_{k+1} > \lambda_k$. Hence, there is no phase transition in the Bayesian information engine. Note that for the naive engine

$$\langle \Delta F_{k+1} \rangle = \alpha \delta_{\mathrm{g}} \int \mathrm{d}\lambda_k \, p(\lambda_k) \int_{\lambda_k}^{\infty} \mathrm{d}y_k \, p(y_k|\lambda_k)(y_k - \lambda_k), \tag{S34}$$
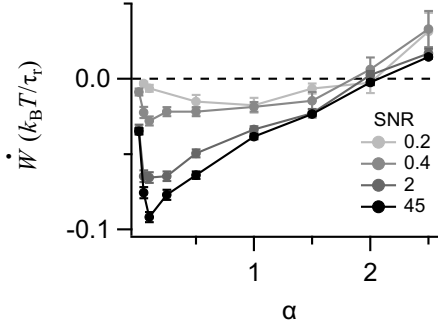
which vanishes for $\alpha = 0$.

FIG. S8. Experiments showing the Bayesian information engine has no phase transition. Trap power $\dot{W}$ as a function of feedback gain α for different signal-to-noise ratios (different shades of gray). Solid curves connect solid markers to ease visualization. Error bars each denote the standard error of the mean.

To corroborate the finding that the Bayesian information engine has no phase transition, we calculate the low-SNR limit of the free-energy gain per time step. When SNR is low, feedback is applied rarely and hence the stationary distribution of bead positions approximates the equilibrium Boltzmann distribution, $p(x_{k+1}|\lambda_k) = \mathcal{N}(x_{k+1}; \lambda_k - \delta_{\mathrm{g}}, 1)$. Hence, the distribution $\hat{p}(\hat{x}_{k+1}|\lambda_k)$ of filter estimates is also approximately Gaussian,

$$\hat{p}(\hat{x}_{k+1}|\lambda_k) = \mathcal{N}\left(\hat{x}_{k+1}; \hat{\mu}, \, \hat{c}\right). \tag{S35}$$

The mean and variance (S11b) of the predictive Kalman filter dictate the steady-state the posterior distribution, of the true bead position x_{k+1} given the filter estimate $\hat{x}_{k+1}$,

$$p(x_{k+1}|\hat{x}_{k+1}) = \mathcal{N}\left(x_{k+1}; \hat{x}_{k+1}, \, \hat{\Sigma}^{\mathrm{SS}}\right), \tag{S36}$$

allowing us to solve for $\hat{p}(\hat{x}_{k+1}|\lambda_k)$,

$$\hat{p}(\hat{x}_{k+1}|\lambda_k) = \int \mathrm{d}u \, p(x_{k+1}|\hat{x}_{k+1} = u) \, \hat{p}(u|\lambda_k), \tag{S37}$$

such that

$$\hat{\mu} = \lambda_k - \delta_{\mathrm{g}} \tag{S38a}$$

$$\hat{c} = 1 - \hat{\Sigma}^{\mathrm{SS}}, \tag{S38b}$$

are the mean and variance of the filter state in Eq. (S37).

Inserting these cumulants into Eq. (S35) and evaluating Eq. (S33b) again using the normalization of $p(\lambda_k)$ gives

$$\langle \Delta F_{k+1} \rangle = \delta_{\mathrm{g}} \sqrt{\frac{2\left(1 - \hat{\Sigma}^{\mathrm{SS}}\right)}{\pi}} \exp\left[-\frac{\delta_{\mathrm{g}}^2}{2\left(1 - \hat{\Sigma}^{\mathrm{SS}}\right)}\right] - \delta_{\mathrm{g}}^2 \left[1 - \mathrm{erf}\,\frac{\delta_{\mathrm{g}}}{\sqrt{2\left(1 - \hat{\Sigma}^{\mathrm{SS}}\right)}}\right]. \tag{S39}$$

Expanding the stationary filter variance (S13) for small $\mathrm{SNR} = 1/\sigma_\mathrm{m}$ yields

$$\hat{\Sigma}^{\mathrm{SS}} = 1 - \frac{\mathrm{SNR}^2}{\mathrm{e}^{2t_\mathrm{s}} - 1} + \mathcal{O}(\mathrm{SNR}^4) \tag{S40}$$

Then, with $\mathrm{erf}(x) \approx 1 - \mathrm{e}^{-x^2}\left(1/x - 1/2x^3\right)/\sqrt{\pi}$ for $x \to \infty$,

$$\langle \Delta F_{k+1} \rangle \approx \sqrt{\frac{2}{\pi}} \left(\frac{\mathrm{SNR}}{\sqrt{\mathrm{e}^{2t_\mathrm{s}} - 1}}\right)^3 \exp\left[-\frac{\delta_\mathrm{g}^2(\mathrm{e}^{2t_\mathrm{s}} - 1)}{2\,\mathrm{SNR}^2}\right] > 0\,, \tag{S41}$$

illustrating that the Bayesian information ratchet extracts free energy at all signal-to-noise ratios.

[1] T. K. Saha, J. N. E. Lucero, J. Ehrich, D. A. Sivak, and J. Bechhoefer, Maximizing power and velocity of an information engine, Proc. Natl. Acad. Sci. U.S.A. **118** (2021).

[2] A. Kumar and J. Bechhoefer, Nanoscale virtual potentials using optical tweezers, Appl. Phys. Lett. **113**, 183702 (2018).

[3] K. Berg-Sørensen and H. Flyvbjerg, Power spectrum analysis for optical tweezers, Rev. Sci. Instrum. **75**, 594 (2004).

[4] J. Bechhoefer, *Control Theory for Physicists* (Cambridge University Press, 2021).

[5] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation* (Prentice Hall, 2000).

[6] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, Springer Series in Synergetics (Springer Berlin Heidelberg, 2010).

[7] Q. Wang and W. E. Moerner, An adaptive anti-Brownian electrokinetic trap with real-time information on single-molecule diffusivity and mobility, ACS Nano **7**, 5792 (2011).

[8] K. Sekimoto, *Stochastic Energetics* (Springer, 2010).