Experimental Observation of the Role of Mutual Information in the Nonequilibrium Dynamics of a Maxwell Demon

J. V. Koski,[1] V. F. Maisi,[1,2,*] T. Sagawa,[3] and J. P. Pekola[1]
[1]Low Temperature Laboratory (OVLL), Aalto University, POB 13500, FI-00076 Aalto, Finland
[2]Centre for Metrology and Accreditation (MIKES), P.O. Box 9, 02151 Espoo, Finland
[3]Department of Basic Science, The University of Tokyo, Komaba 3-8-1, Meguro-ku, Tokyo 153-8902, Japan
(Received 6 May 2014; revised manuscript received 18 June 2014; published 14 July 2014)

We validate experimentally a fluctuation relation known as generalized Jarzynski equality governing the work distribution in a feedback-controlled system. The feedback control is performed on a single electron box analogously to the original Szilard engine. In the generalized Jarzynski equality, mutual information is treated on an equal footing with the thermodynamic work. Our measurements provide the first evidence of the role of mutual information in the fluctuation theorem and thermodynamics of irreversible processes.

DOI: 10.1103/PhysRevLett.113.030601 PACS numbers: 05.70.Ln, 65.40.gd, 72.15.Eb, 73.23.Hk

The second law of thermodynamics gives the inevitable upper bound of the available work that we can extract from fuels and heat baths. Information has been recognized as another kind of thermodynamic fuel that can be used to extract work with measurement and feedback control. The relation between information and thermodynamics is a topic of long-standing interest in the field of statistical physics, dating back to the thought experiment of a "Maxwell demon" [1–3]. Relatively recent progress with universal nonequilibrium equalities applying to irreversible processes, known as fluctuation theorems [4–11], has brought renewed attention to this problem. In particular, the second law of thermodynamics and nonequilibrium equalities have been generalized to irreversible processes that involve information treatment, such as measurement, feedback control, or information erasure [12–23].

A Maxwell demon is an object that measures the microscopic state of a system and drives it to extract work or store energy with the aid of the measurement outcome. A crucial element for the fidelity of this operation is mutual information $\langle I\rangle$. It characterizes the correlation between the state of the measured thermodynamic system and the measurement outcome stored into the memory of the measurement device, and as such describes the efficiency of the measurement. Several recent experiments have illustrated the relation between information and thermodynamics [24–27]; however, none have yet demonstrated the role of mutual information in irreversible feedback processes.

In this Letter, we study experimentally the mutual information in a feedback-controlled device and provide the first demonstration of its connection to the fluctuation theorem. When the state of a generic thermodynamic system in state n is measured with an outcome m, the stochastic mutual information [17,22] is defined as

$$I(m,n) \coloneqq \ln P(n|m) - \ln P(n), \tag{1}$$

where $P(n)$ is the initial probability of the state being n, whereas $P(n|m)$ is the probability that it is n under the condition that the measurement outcome is m. As $I(m,n)$ depends on the probability distribution of (m,n), we need to measure many samples in the ensemble in order to determine the value of $I(m,n)$, as is the case for stochastic Shannon entropy [28]. Jarzynski equality (JE) [5],

$$\langle e^{-(W-\Delta F)/k_B T}\rangle = 1, \tag{2}$$

has been generalized to systems with measurement and feedback control [17] to

$$\langle e^{-(W-\Delta F)/k_B T - I}\rangle = 1, \tag{3}$$

where W is the applied work, ΔF is the change in free energy, T is the temperature of the thermal reservoir, and k_B is the Boltzmann constant. Equation (3) further reproduces the second law of thermodynamics as

$$\langle W\rangle - \Delta F \geq -k_B T\langle I\rangle, \tag{4}$$

where mutual information [29] $\langle I\rangle$ is the expectation value of the stochastic mutual information. As $\langle I\rangle$ is maximized in the ideal limit of the measurement correlating perfectly with the actual state, i.e., $P(n|m) = \delta_{mn}$, the magnitude of $\langle I\rangle$ describes the efficiency of the measurement, providing the upper limit to how much work can be extracted from the system for the given information. We further define

$$\eta_f \coloneqq \frac{-(\langle W\rangle - \Delta F)}{k_B T\langle I\rangle} \leq 1 \tag{5}$$

to describe the efficiency of the feedback control. If $\eta_f = 1$, the feedback control is perfect and thermodynamically reversible, where all of the mutual information is extracted as work. The condition to achieve the reversible feedback has been discussed in Ref. [20].

We perform the following experiment in a feedback-controlled two-state system. Our device is a single-electron box [30,31] (SEB), illustrated in Fig. 1(a), which connects two metallic islands by a junction, permitting electron transport between the two by tunneling. We employ two-island configuration in our box [27,32,33]: one island is made out of copper, which is a normal metal, and the other out of aluminum, which is a superconductor. The superconductor energy gap in the density of states strongly suppresses the tunneling rates to observable levels. Furthermore, as both islands have only capacitive coupling to the environment, the electric noise to the SEB is minimal. The SEB is placed in a dilution cryostat, and the experiments are performed at $T = 100 \pm 3$ mK. The two islands have a mutual capacitance C_Σ, such that tunneling electrons change the charge of this capacitor by elemetary charge $-e$ per electron. The charging energy of an SEB is

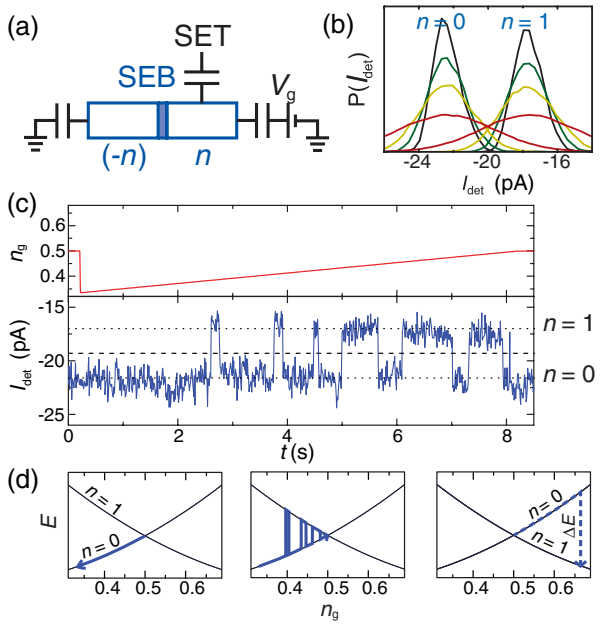

FIG. 1 (color online). Device and operation. (a) The single electron box (SEB), highlighted in blue, is the system under study. It is monitored by a single electron transistor (SET), whose current $I_{\rm det}$ depends on n, the number of excess electrons on the right island of the SEB. The SEB is controlled by gate voltage V_g. (b) Single trace histogram of detector signal for states $n = 0$ (peaks to the left) and $n = 1$ (peaks to the right) with filter cutoff frequencies 50 (black), 100 (green), 300 (yellow), and 1000 Hz (red), in the order of decreasing maxima. (c) A full trace of the feedback control. $I_{\rm det}$ shows the measured occupation in the SEB. (d) Energy diagrams of the process. The rapid feedback (left) extracts work by lowering the energy. During the return back to degeneracy (middle), net work is extracted from the thermal excitations of n entering the higher energy state. If the rapid feedback were performed incorrectly (right), excess work equal to ΔE would be applied to add energy to the system. The return to degeneracy would again follow the behavior of middle panel.

$$E(n, n_g) = E_C(n - n_g)^2, \tag{6}$$

where $E_C = e^2/2C_\Sigma$ is the energy required to charge the capacitor by a single electron, and $-en$ is the charge of the right island, induced by n electrons that have tunneled from the left island. Our SEB has $E_C \approx 111$ μeV. Consequently, charge conservation requires that the charge of the left island is en. The electron tunneling is controlled by a nearby gate, accumulating a charge equal to $en_g = C_g V_g$ to the gate capacitor. The gate voltage V_g is modulated to drive the SEB with n being the stochastic state that changes by electron tunneling. The state n naturally favors the energy minimum given by Eq. (6), but can also change to a higher energy state due to thermal excitations. The islands of the SEB are a few μm long, providing a sufficiently small C_Σ at sub-Kelvin temperatures to achieve $E_C \gg k_B T$. Then the SEB is a two-state system with either $n = 0$ or $n = 1$ if we operate in the range $n_g = 0 \ldots 1$. A nearby single electron transistor (SET) monitors n. The measured trajectories of n then determine the applied work $W = \int dt (dn_g/dt)(\partial E/\partial n_g)$.

An SEB can be driven and monitored to test thermodynamic relations in a two-state system [34], and has already been used to verify various fluctuation relations [32,33]. It can also be operated [27] as a Szilard engine [2], ideally extracting $k_{\rm B} T \ln 2$ of work per feedback cycle. The steps of the operation follow the description introduced in Ref. [20]. The initial energies of states $n = 0$ and $n = 1$ are equal by setting $n_g = 0.5$. Then n follows the distribution with equal probabilities $P(0) = P(1) = 1/2$. The state n is measured with the SET, providing an outcome m. As feedback control, the gate is rapidly driven to $n_g = 0.5 \pm \Delta n_g$, where Δn_g is a predetermined parameter set to $\Delta n_g = 0.167$ for the present experiment, $+$ sign is used for $m = 1$, and $-$ sign for $m = 0$. This drive causes the state m to have lower energy by $\Delta E = 2E_C \Delta n_g$ than the other state. Finally, n_g is slowly brought back to degeneracy $n_g = 0.5$, extracting net work from concurrent thermal excitations of n. In this closed cycle, the free energy difference over the whole cycle is zero, $\Delta F = 0$, and we only need to consider W.

Let ε be the error rate of the measurement, which is assumed to be equal for measuring $n = 0$ and $n = 1$; we obtain an incorrect outcome with probability $P(n|m) = \varepsilon$ for $m \neq n$, while $P(n|m) = 1 - \varepsilon$ for $m = n$. If $\varepsilon = 0$, the measurement is error free and $n = m$ holds. By direct insertion to Eq. (1), we obtain stochastic mutual information $I(n, m) = \ln[2(1 - \varepsilon)]$ for $m = n$, and $I(n, m) = \ln(2\varepsilon)$ for $m \neq n$. The average of I over all possible n and m produces the mutual information:

$$\begin{aligned} \langle I \rangle &= \sum_{nm} P(n, m) I(n, m) \\ &= \ln 2 + (1 - \varepsilon) \ln(1 - \varepsilon) + \varepsilon \ln \varepsilon, \end{aligned} \tag{7}$$

which is the difference between the unconditional Shannon entropy $-\sum_n P(n)\ln P(n)$ and the conditional one $-\sum_{nm} P(n|m)\ln P(n|m)$. The mutual information is nonnegative and not greater than the Shannon information of the outcome; $\langle I\rangle \leq \ln 2$ holds in our setup, where the equality is achieved if the measurement is error free ($\varepsilon = 0$). The mutual information created by the measurement can be used to implement feedback control that enables us to extract useful work.

To introduce the measurement error in the experiment, we change the cutoff frequency of the numerical low-pass filter applied to the detector signal for each measurement. The signal for reading n is subject to noise; increasing the cutoff frequency enhances noise in the filtered readout, as shown in Fig. 1(b). This way, the probability ε to measure the state n incorrectly is different for each cutoff frequency. In the analysis, for determining the trajectory of n at the time of the measurement and during the foregoing feedback control, we filter the data with a low cutoff frequency (50 Hz), such that signal-to-noise ratio is high, and apply threshold detection as in Fig. 1(c). We assume that the obtained n is the true value of the charge state: the histogram of Fig. 1(b) shows that the signal overlap for 0 and 1 is small. We estimate the approximate error for n to be below 0.2%. The tunneling rate at degeneracy, $\Gamma_0 = 2.7$ Hz, remains lower than the cutoff frequency of the filter and thus all the relevant transitions of n are detected. The error probability ε is extracted by counting the number of process cycles, where $n \neq m$.

Ideally, n_g is driven instantly after the measurement to obtain the energy difference ΔE between the states such that the initial state is at energy minimum. The work done in the fast drive is determined by the energy difference of Eq. (6), $W_0 = E(n, n_g \pm \Delta n_g) - E(n, n_g)$. After the fast response, the consequent slow return back to degeneracy practically starts from thermal equilibrium, as the rate of equilibration is significantly faster than the rate of the drive. Let $P_0(W)$ be the probability distribution for applied work W for an ideal fast feedback response followed by slow return to degeneracy. The slow drive satisfies JE (2), and since the fast feedback response produces a fixed W_0, the distribution satisfies $\langle e^{-W/k_BT}\rangle_0 = e^{-(\Delta F_0 + W_0)/k_BT} = 2/(1 + e^{-\Delta E/k_BT})$, where ΔF_0 is the free energy difference over the drive back to degeneracy, and $\langle \ldots\rangle_0$ denotes averaging over $P_0(W)$. This condition allows us to determine the extracted work as $-\langle W\rangle_0 = rk_BT\ln[2/(1 + e^{-\Delta E/k_BT})]$, where r is determined by the drive rate and has a value between 0 and 1.

In the case of an incorrect feedback response, an additional ΔE work is paid as illustrated in Fig. 1(d), and work distribution $P_0(W - \Delta E)$ is followed. This occurs either by a measurement error, or by a finite delay τ between initiating the measurement and triggering the feedback. During this delay, the distribution of the states of the SEB evolves naturally, and the probability for the state to be different by the time the feedback takes place is $\delta = \frac{1}{2}(1 - e^{-2\Gamma_0\tau})$. In the presented experiment, $\tau \simeq 15$ ms. On the other hand, if measurement is incorrect *and* the state changes after the measurement, the distribution again follows $P_0(W)$, and we obtain

$$P(W) = [(1-\varepsilon)(1-\delta) + \varepsilon\delta]P_0(W) + [\varepsilon(1-\delta) + (1-\varepsilon)\delta]P_0(W - \Delta E), \quad (8)$$

matching the measured distributions shown in Figs. 2(a)–2(c). Incidences with correct and incorrect measurement results have different values for I, and the resulting distribution modified by mutual information is

$$\begin{aligned} P(\bar{W} \equiv W + k_BTI) &= (1-\varepsilon)(1-\delta)P_0\{\bar{W} - k_BT\ln[2(1-\varepsilon)]\} \\ &+ \varepsilon(1-\delta)P_0[\bar{W} - \Delta E - k_BT\ln(2\varepsilon)] \\ &+ (1-\varepsilon)\delta P_0\{\bar{W} - \Delta E - k_BT\ln[2(1-\varepsilon)]\} \\ &+ \varepsilon\delta P_0[\bar{W} - k_BT\ln(2\varepsilon)], \end{aligned} \quad (9)$$

which follows the generalized JE (3). The measured distributions shown in Figs. 2(d)–2(f) match Eq. (9). In Fig. 2(f), the four peaks in the distribution are numbered in the order listed in Eq. (9).

The average extracted work given by the distribution of Eq. (8) is

$$-\langle W\rangle = rk_BT\ln\left(\frac{2}{1 + e^{-\Delta E/k_BT}}\right) - \varepsilon_F\Delta E, \quad (10)$$

where $\varepsilon_F = \varepsilon(1-\delta) + (1-\varepsilon)\delta$ is the probability for incorrect feedback. The extracted work is maximized by setting Δn_g such that

$$\Delta E/k_BT = \ln(r/\varepsilon_F - 1), \quad (11)$$

with which, in the limits of $r \to 1$ and $\tau \to 0$, Eq. (10) becomes an equality with Eq. (4), as has been demonstrated in [20]. For any other ΔE, r or τ, the extracted work is smaller in agreement with the second law of thermodynamics. In the ideal limit of $r \to 1$, $\varepsilon \to 0$, $\tau \to 0$, and correspondingly, $\Delta E \to +\infty$, we obtain $-\langle W\rangle \to k_BT\ln 2$ as is the case for the conventional Szilard engine.

The generalized JE has also another form, $\langle e^{-(W-\Delta F)/k_BT}\rangle = \gamma$ [17]. Here, γ is a parameter that quantifies the efficiency of both the measurement and feedback. While this equality has been verified in a colloidal system [24], the present Letter is the first test of the generalized JE that connects thermodynamics and the mutual information, Eq. (3). The distribution of Eq. (8) produces

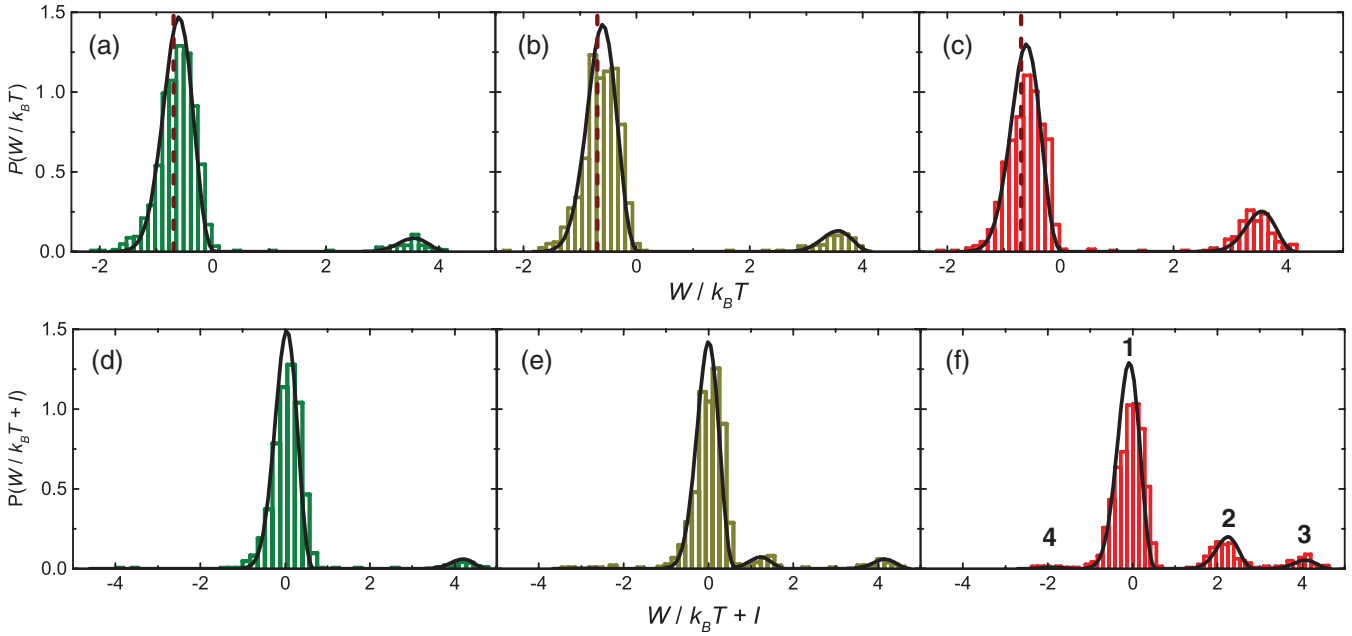

FIG. 2 (color online). Distributions of work with mutual information. [(a)–(c)] The measured W/k_BT for different measurement error probabilities $\varepsilon = 0.02$, 0.05, and 0.13 with $N = 1177$, 925, and 1025, respectively. Black lines show the numerically obtained distributions. [(d)–(f)] The distributions with the mutual information I added to W/k_BT. The numbers in the panel (f) refer to the four peaks in Eq. (9).

$$\langle e^{-W/k_BT}\rangle = \frac{2}{1+e^{-\Delta E/k_BT}} - 2\varepsilon_F \tanh\left(\frac{\Delta E}{2k_BT}\right). \quad (12)$$

Figure 3 shows the measured expectation values discussed above as a function of measurement error. As one approaches the low-error regime, the incidences of incorrect feedback response become increasingly rare, and the average extracted work tends to approximately $0.7k_BT\ln(2)$, as shown in Fig. 3(a). The feedback efficiency η_f, given by Eq. (5), remains almost constant for the lowest ε, and the extracted work primarily depends on measurement efficiency. For higher ε, the feedback protocol should be changed for a better η_f. The protocol could be optimized by correspondingly reducing the applied energy difference ΔE in accordance with Eq. (11). Figure 3(b) shows the results for the test of the generalized JE. We see that Eq. (3) remains valid within measurement errors.

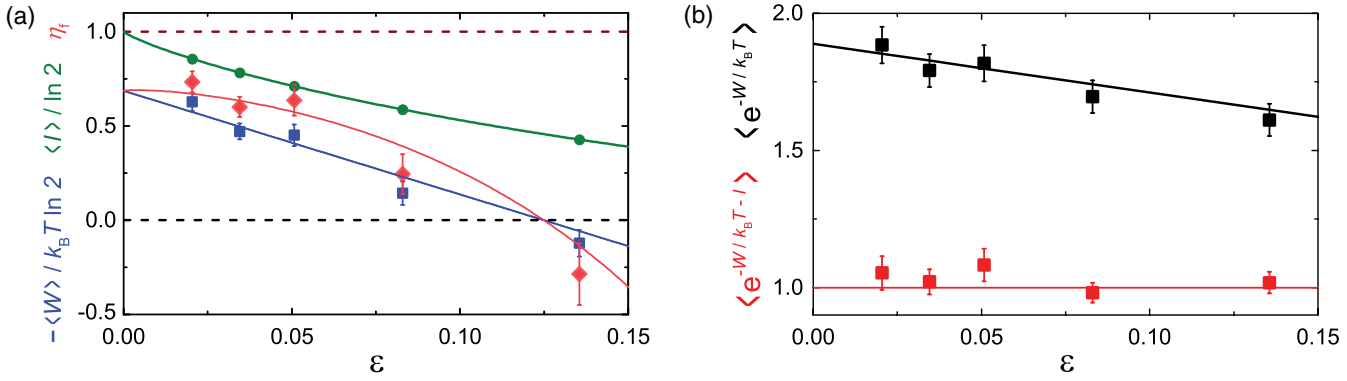

FIG. 3 (color online). Efficiency and fluctuation relations. (a) Average extracted work (blue squares and solid line) and the feedback efficiency (red diamonds and solid line) as functions of measurement error. Experimental data are shown by symbols (in both panels), mutual information obtained from Eq. (7) is shown by the green circles and solid line, and numerical predictions for other quantities by solid lines (in both panels). The dashed brown line shows the fundamental maximum for η_f, $\langle I\rangle$, and $-\langle W\rangle$. The black dashed line shows the limit for η_f and $-\langle W\rangle$ below which the process is dissipative. The results are obtained from $N = 1177$, 1734, 925, 1060, and 1025 repetitions, in the order of increasing ε. (b) Test of the generalized JE with mutual information. The error bars include the statistical error, as well as the uncertainity in the measured value of E_C/k_BT.

In conclusion, our experiments illustrate the role of mutual information in the performance of a Maxwell demon. We show that our device follows generalized Jarzynski equality when under feedback control similar to that of a Szilard engine. With a fixed feedback protocol, we show that the efficiency of the feedback changes with the measurement efficiency.

This work has been supported in part by Academy of Finland (Project No. 139172) and its LTQ (Project No. 250280), the National Doctoral Programme in Nanoscience, NGS-NANO (V.F.M.), the European Union Seventh Framework Programme INFERNOS (FP7/2007-2013) under Grant Agreement No. 308850, and JSPS KAKENHI Grants No. 25800217 and No. 22340114 (T. S.). We thank O.-P. Saira for useful discussions. We acknowledge OMN, Micronova Nanofabrication Centre, and the Cryohall of Aalto University for providing the processing facilities and technical support.

*Present address: Solid State Physics Laboratory, ETH Zürich, 8093 Zürich, Switzerland.

[1] J. C. Maxwell, *Theory of Heat* (Appleton, London, 1871).
[2] L. Szilard, Z. Phys. **53**, 840 (1929).
[3] *Maxwell's Demon*, edited by H. S. Leff and A. F. Rex (IOP Publishing, Bristol, 2003).
[4] G. N. Bochkov and I. E. Kuzovlev, Sov. Phys. JETP **45**, 125 (1977).
[5] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[6] J. Kurchan, J. Phys. A **31**, 3719 (1998).
[7] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).
[8] C. Jarzynski, J. Stat. Phys. **98**, 77 (2000).
[9] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).
[10] M. Esposito, U. Harbola, and S. Mukamel, Rev. Mod. Phys. **81**, 1665 (2009).
[11] M. Campisi, P. Hänggi, and P. Talkner, Rev. Mod. Phys. **83**, 771 (2011).
[12] K. Shizume, Phys. Rev. E **52**, 3495 (1995).
[13] B. Piechocinska, Phys. Rev. A **61**, 062314 (2000).
[14] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. Lett. **98**, 080602 (2007).
[15] M. Esposito and C. Van den Broeck, Europhys. Lett. **95**, 40004 (2011).
[16] H. Touchette and S. Lloyd, Phys. Rev. Lett. **84**, 1156 (2000).
[17] T. Sagawa and M. Ueda, Phys. Rev. Lett. **104**, 090602 (2010).
[18] T. Sagawa and M. Ueda, New J. Phys. **15**, 125012 (2013).
[19] J. M. Horowitz and S. Vaikuntanathan, Phys. Rev. E **82**, 061120 (2010).
[20] J. M. Horowitz and J. M. R. Parrondo, Europhys. Lett. **95**, 10005 (2011).
[21] Y. Morikuni and H. Tasaki, J. Stat. Phys. **143**, 1 (2011).
[22] D. Abreu and U. Seifert, Phys. Rev. Lett. **108**, 030601 (2012).
[23] S. Deffner and C. Jarzynski, Phys. Rev. X **3**, 041003 (2013).
[24] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano, Nat. Phys. **6**, 988 (2010).
[25] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, Nature (London) **483**, 187 (2012).
[26] A. Bérut, A. Petrosyan, and S. Ciliberto, Europhys. Lett. **103**, 60002 (2013).
[27] J. V. Koski, V. F. Maisi, J. P. Pekola, and D. V. Averin, arXiv:1402.5907
[28] G. Cuetara, M. Esposito, and A. Imparato, Phys. Rev. E **89**, 052119 (2014).
[29] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley and Sons, New York, 1991).
[30] M. Büttiker, Phys. Rev. B **36**, 3548 (1987).
[31] P. Lafarge, H. Pothier, E. R. Williams, D. Esteve, C. Urbina, and M. H. Devoret, Z. Phys. B **85**, 327 (1991).
[32] O.-P. Saira, Y. Yoon, T. Tanttu, M. Möttönen, D. V. Averin, and J. P. Pekola, Phys. Rev. Lett. **109**, 180601 (2012).
[33] J. V. Koski, T. Sagawa, O-P. Saira, Y. Yoon, A. Kutvonen, P. Solinas, M. Möttönen, T. Ala-Nissila, and J. P. Pekola, Nat. Phys. **9**, 644 (2013).
[34] D. V. Averin and J. P. Pekola, Europhys. Lett. **96**, 67004 (2011).