

## Influence of measurement error on Maxwell's demon

Vegard Sørdal,<sup>1</sup> Joakim Bergli,<sup>1</sup> and Y. M. Galperin<sup>1,2</sup>

<sup>1</sup>*Department of Physics, University of Oslo, 0316 Oslo, Norway*

<sup>2</sup>*A. F. Ioffe Physico-Technical Institute of Russian Academy of Sciences, 194021 St. Petersburg, Russia*

(Received 17 January 2017; revised manuscript received 29 March 2017; published 21 June 2017)

In any general cycle of measurement, feedback, and erasure, the measurement will reduce the entropy of the system when information about the state is obtained, while erasure, according to Landauer's principle, is accompanied by a corresponding increase in entropy due to the compression of logical and physical phase space. The total process can in principle be fully reversible. A measurement error reduces the information obtained and the entropy decrease in the system. The erasure still gives the same increase in entropy, and the total process is irreversible. Another consequence of measurement error is that a bad feedback is applied, which further increases the entropy production if the proper protocol adapted to the expected error rate is not applied. We consider the effect of measurement error on a realistic single-electron box Szilard engine, and we find the optimal protocol for the cycle as a function of the desired power  $P$  and error  $\varepsilon$ .

DOI: [10.1103/PhysRevE.95.062129](https://doi.org/10.1103/PhysRevE.95.062129)

### I. INTRODUCTION

Maxwell's demon was introduced as a thought experiment to illustrate the statistical nature of the second law of thermodynamics [1]. The demon has very sharp powers of observation, so it can detect the motion of individual molecules. In addition, it can rapidly act on the basis of its observations and thereby sort fast and slow molecules. This makes heat flow from the cold to the hot side, apparently without the need for any work, in contradiction to the second law of thermodynamics. For some time it was thought that the act of observation necessarily required some amount of work [2,3]. The present consensus [4,5] seems to be that the observation, in principle, can be performed without work. At the same time, the erasure of the information obtained, being a logically irreversible operation, also is thermodynamically irreversible and has a necessary cost in terms of work that is converted to heat. However, there is still some controversy on this point [6–8].

Modern technology now enables us to be as accurate in observation and quick in action as the imagined demon. Recently, several experiments, in which close analogies to the original thought experiment were realized, have been reported in a range of physical systems: atoms [9–11], colloidal particles [12,13], molecules [14], electrons [15–17], and photons [18]. This shift from imagined to real experiments motivates us to study the impact of measurement errors on the performance of experimental Maxwell's demons.

If there is some chance that the measurement result is wrong, it means that the correlation between the state of the system and the measurement device is not perfect. That is, the mutual information between the two is less than the full information of the logical states of the measurement device. In [8], Sagawa and Ueda show that the traditional Landauer bound  $W \geq T \ln 2$  (we use units where the Boltzmann constant  $k_B = 1$ ) only holds for a symmetric memory, and the total work expended on measurement and memory erasure has a lower bound given by the mutual information  $I$  between the system and the measurement device,

$$W_{\text{measure}} + W_{\text{erase}} \geq TI. \quad (1)$$

The right-hand side is exactly the same as the heat that can be extracted from a thermal bath using the information about

the system. Although measurement errors will give reduced mutual information, we argue that it will not be possible to reach equality in Eq. (1) in this case. To justify this, consider the extreme case in which the mutual information  $I$  is zero, i.e., there is a 50% chance that the measurement is wrong. In this case, the measurement can be done reversibly without any work, but there will still be one bit of information stored in the memory that has to be erased with a cost of  $T \ln 2$  according to Landauer.

### II. ANALYSIS OF A MODEL SYSTEM

To clearly show the difference between a true measurement error and a process that saturates Eq. (1), we will analyze a simple model. Consider a total system (memory + system) with a phase space  $\mathcal{P}$ . We divide its phase space in subspaces  $\mathcal{P}_i$ , each of which corresponds to a specific logical state. With the probability distribution of the total phase space denoted  $P(x)$ , the probability distribution of the logical states is

$$P_L(i) = \sum_{x \in \mathcal{P}_i} P(x) \quad (2)$$

and the conditional probability of the microstate  $x$  given the logical state  $i$  is

$$P(x|i) = P(x)/P_L(i). \quad (3)$$

The total entropy  $S$ , logical entropy (information)  $H$ , and conditional entropy  $S(\mathcal{P}_i|i)$  are then given by

$$S = - \sum_x P(x) \ln P(x), \quad H = - \sum_i P_L(i) \ln P_L(i),$$

$$S(\mathcal{P}_i|i) = - \sum_{x \in \mathcal{P}_i} P(x|i) \ln P(x|i). \quad (4)$$

The conditional entropy can be thought of as the internal physical entropy of the distribution  $P(x|i)$  on  $\mathcal{P}_i$  for each of the logical states  $i$ . The average conditional entropy is  $S_{\text{in}} = \sum_i P_L(i) S(\mathcal{P}_i|i)$ , which we call the internal entropy. It follows that we can write the total entropy as a sum,

$$S = H + S_{\text{in}}. \quad (5)$$

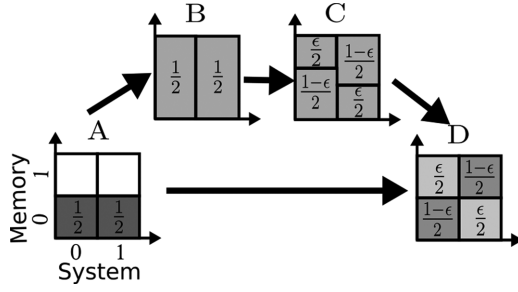


FIG. 1. Schematic of reversible and irreversible measurement in a two-bit total system (system + memory). The four logical subspaces are 00/01/10/11.

With this formalism, we can analyze the model system shown in Fig. 1. The system is a standard Szilard engine, with a single molecule in a box with a dividing wall that can be inserted, removed, and used as a piston. The memory is represented by an equivalent single-molecule box. Consequently, we have four logical states. The phase space of each molecule is reduced to one dimension by only considering the movement of the molecule in the direction in which the volume of the compartments expands and contracts, and ignoring the momentum, as all processes will be isothermal and therefore the momentum distribution is independent of the protocol. The relevant part of the total phase space is then two-dimensional, and we represent the position of the molecule in the system on the horizontal axis, and in the memory on the vertical axis. To calculate the total entropy, we use Eq. (5) and the fact that the conditional entropy of a system uniformly distributed in a given region of phase space is given by the logarithm of the phase-space volume, which we show in the following subsection.

### A. Conditional entropy

The free-energy of an ideal gas in a three-dimensional box is

$$F(T, V, N) = -NT \ln \left[ \frac{V e}{N} \left( \frac{mT}{2\pi\hbar^2} \right)^{3/2} \right], \quad (6)$$

which we use to calculate the entropy,

$$\begin{aligned} S &= -\left( \frac{\partial F}{\partial T} \right)_{V, N} = N \left\{ \frac{3}{2} + \ln \left[ \frac{eV\sqrt{2}}{4N} \left( \frac{mT}{\pi\hbar^2} \right)^{3/2} \right] \right\} \\ &= N \left[ \frac{3}{2} + \ln \left( \frac{V}{NV_q} \right) \right], \quad \frac{1}{V_q(T)} \equiv \frac{e\sqrt{2}}{4} \left( \frac{mT}{\pi\hbar^2} \right)^{3/2}. \end{aligned} \quad (7)$$

Here  $V_q$  is of the order of the de Broglie wavelength. To keep the classical limit, we have to assume that  $V \gg NV_q$ . Further, we will deal with one particle. Therefore,

$$S = \frac{3}{2} + \ln \left( \frac{V}{V_q} \right). \quad (8)$$

Our system consists of two volumes (system and memory) with one particle in each, and we need to calculate the conditional entropy for the configurations  $A$  and  $D$  of Fig. 1 (hereafter denoted  $1A$  and  $1D$ ). In each of the four logical states,  $i = 00, 01, 10, 11$ , the internal states of the system and memory

are uncorrelated, and the conditional entropy is the sum of two contributions of the type shown in Eq. (8). Denoting the position of the dividing wall in the system as  $x_S$  and in the memory  $x_M$  and the length of the box with the gas  $L$ , the conditional entropy in the logical state  $i$  can be specified as

$$\begin{aligned} S(\mathcal{P}_i|i) &= S_0 + \ln \left( \frac{x_S x_M}{L^2} \right), \\ S_0 &\equiv 3 + 2 \ln \left( \frac{V}{V_q} \right) \gg 1. \end{aligned} \quad (9)$$

In the following, we omit the constant  $S_0$ , which means that the entropy of the reference state where the dividing walls are removed is zero, and all given entropies are entropy changes from this reference state.

### B. Irreversible measurement

We perform a measurement on the system and store it in the memory. Throughout the paper, we assume that the measurement is classical and does not affect the state of the system. If there is a probability  $\varepsilon$  that the measurement gives the wrong result, we have a transition from  $A$  to  $D$  in Fig. 1. For configuration  $1A$ , we have two states with probabilities  $\frac{1}{2}$  each, giving

$$H^A = - \sum_i P_L(i) \ln P_L(i) = -2 \times \frac{1}{2} \ln \frac{1}{2} = \ln 2. \quad (10)$$

Using Eq. (9) with  $x_S = x_M = L/2$  and omitting  $S_0$ , we get for the conditional entropy  $S(\mathcal{P}_{00}|00) = S(\mathcal{P}_{10}|10) = \ln \frac{1}{4}$  and the internal entropy

$$S_{\text{in}}^A = \sum_i P_L(i) S(\mathcal{P}_i|i) = 2 \times \frac{1}{2} \ln \frac{1}{4} = -2 \ln 2. \quad (11)$$

Consequently, the total entropy of configuration  $1A$  is

$$S^A = H^A + S_{\text{in}}^A = -\ln 2. \quad (12)$$

For configuration  $1D$ , we obtain in a similar way

$$H^D = -2 \times \frac{\varepsilon}{2} \ln \frac{\varepsilon}{2} - 2 \times \frac{1-\varepsilon}{2} \ln \frac{1-\varepsilon}{2} = \ln 2 + S_\varepsilon, \quad (13)$$

$$S_{\text{in}}^D = 2 \times \frac{\varepsilon}{2} \ln \frac{1}{4} + 2 \times \frac{1-\varepsilon}{2} \ln \frac{1}{4} = -2 \ln 2, \quad (14)$$

$$S^D = H^D + S_{\text{in}}^D = -\ln 2 + S_\varepsilon, \quad (15)$$

where  $S_\varepsilon \equiv -\varepsilon \ln \varepsilon - (1-\varepsilon) \ln(1-\varepsilon)$ . The negative-valued entropies are due to the omitted constant  $S_0$ . We see that the total entropy in the transition from  $1A$  to  $1D$  is irreversibly increased by an amount  $S^D - S^A = S_\varepsilon$ . Since both the system and the memory have equal probabilities of being in their two logical states, the logical information in each is  $H_{\text{system}}^D = H_{\text{memory}}^D = \ln 2$ . The mutual information between the system and memory is

$$I^D = H_{\text{system}}^D + H_{\text{memory}}^D - H^D = \ln 2 - S_\varepsilon.$$

**C. Reversible measurement**

The transition from configuration 1A to 1D can also be achieved reversibly while extracting work if we consider the following steps (this process is also considered in [19]):

1A → 1B. In the transition from 1A to 1B we isothermally expand the state 0 of the memory. This allows the particle to expand into the full volume of the memory. In this process, work  $W$  is performed by the system, and heat  $Q = W$  is taken from the reservoir. The entropy change is

$$\Delta S = W/T = \ln 2$$

with a corresponding entropy decrease in the reservoir.

1B → 1C. We then perform a measurement on the system, and we reinsert the partition wall in the memory according to the result obtained. There is no error in this measurement, and the correlation between the position of the dividing wall of the memory and the position (left or right) of the gas molecule of the system is perfect. Here  $\varepsilon$  is just a parameter that describes where we insert the divider in the memory. There is no entropy change.

1C → 1D. We then compress the divider of the memory isothermally back to the central position. In this process, we have to perform work on the system, but an amount less than the work performed by it in the transition from 1A to 1B. The entropy change is

$$\Delta S = W/T = S_\varepsilon - \ln 2.$$

In our view, this process does not represent a real measurement error, which is irreversible and has an associated entropy production  $S_\varepsilon$  due to Gibbs or environmental coarse-graining. The final state of this process (1D) is the same as the one obtained when there was a measurement error, but the whole process is thermodynamically reversible, and the reduction of the environmental entropy is exactly the same as the increase of the system entropy. In the process, we have extracted the net work from the thermal bath, so that the work of measurement that enters Eq. (1) is  $W_{\text{measure}} = -TS_\varepsilon$ , which is negative. Erasing the memory requires  $W_{\text{erase}} = T \ln 2$  according to the usual Landauer's principle, which gives

$$W_{\text{measure}} + W_{\text{erase}} = T \ln 2 - TS_\varepsilon = TI^D,$$

which saturates the inequality (1).

**D. Origin of the irreversible measurement entropy**

To get a deeper understanding of the irreversible nature of a measurement with error, consider Fig. 2. In Fig. 2, step A (hereafter 2A), we have the same initial state as before. Figure 2, step B (hereafter 2B) shows the state just after the measurement was performed. Most of the initial states in the phase space are mapped to the correct final region, but a small

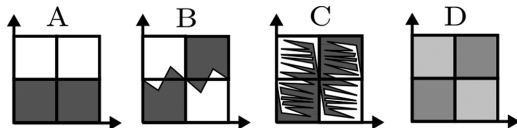


FIG. 2. How a system evolves from step A to D in Fig. 1 after a measurement error.

fraction gets mapped to a different region. This corresponds to the cases in which the result of the measurement does not agree with the actual position of the system molecule. If the system and the measurement device constitute an isolated system during the operation, and no other degrees of freedom are involved, the mapping from 2A to 2B would be described by a deterministic Hamiltonian evolution in time. Liouville's theorem then guarantees that the entropy of the final state is the same as that in the initial state. If the evolution is affected by other microscopic degrees of freedom in the device or the environment, which is certainly realistic in most cases, the mapping would be stochastic, and it depends on these additional degrees of freedom. We assume that after 2B, the phase points will never again cross the lines separating the different logical states. The physical meaning of this is that the barriers between the states are infinite. In a short time, the phase-space region where the system can be found will develop into some complicated shape 2C, but for a closed system the entropy will still be the same. Now we have to appeal to some coarse-graining procedure. For a closed system, we refer to the phase-space coarse-graining introduced by Gibbs (see Ref. [20] for a recent discussion). In the presence of some interaction with an environment, coarse-graining occurs over dynamical evolution [21,22]. In this way, the complex structure of the accessible phase space in 2C is rendered indistinguishable and is replaced by the uniform distribution in 2A. This step is irreversible and increases the total entropy of the system by  $S_\varepsilon$  without any decrease in entropy anywhere.

**III. MODEL FOR A SZILARD ENGINE WITH MEASUREMENT ERRORS**

To study the effects of this measurement error, we will now analyze a model of an experimentally realized Szilard engine [15]. This model is comprised of a single-electron box consisting of two metallic islands connected by a tunnel junction. The existence of an additional electron on one of the two islands can be measured by the charge configuration of the box, and its state can be controlled by gate voltages applied to the islands, giving a time-dependent potential difference  $V(t)$  between the two islands. Work can be extracted from the system via the following procedure:

- (i) Perform a measurement, and quickly set the potential of the occupied island to zero while raising the potential of the empty island to some value  $V_0 \equiv V(0^+)$ .
- (ii) Reduce the potential of the raised island according to some protocol  $V(t)$  until time  $t = \tau$ , at which point we start over from step (i).

There is a probability that the electron will tunnel between the two islands, and whenever the electron occupies the island where the potential is being decreased, heat is extracted from the environment and converted to work. We imagine that we are continuously repeating steps (i) and (ii) above, and we want to minimize the total entropy production rate when varying the driving protocol  $V(t)$  and the time  $\tau$ . In the experimental protocol in [15], the potential difference was always reduced back to zero before the next measurement. In our optimal protocol, it does not need to be zero since the energy gained from tunneling decreases with the potential difference, and longer  $\tau$  gives smaller power.

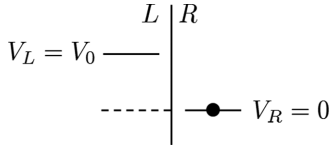


FIG. 3. Level diagram after the first measurement and raising the potential of the empty island that does not cost any work.

### Details of the protocol

We have a set of two islands—left ( $L$ ) and right ( $R$ )—divided by a tunnel barrier. The islands are gated, so their potentials can be manipulated independently. The pair of islands contains only one excess electron, so each island may contain either zero or one excess electron; the occupancy of each island can be measured, say, by single electron transistor. A similar setup was used in the experiment [15]. Let us start the protocol from the state where we have measured an electron on the right island. Then we quickly raise the potential of the empty island to the value  $V_0$ . After this procedure, we arrive at the level diagram shown in Fig. 3. Then we decrease the potential of the raised island according to some protocol  $V(t)$  in the time  $t = [0, \tau]$ , as shown in Fig. 4 (left panel). At time  $t = \tau$ , we reach the situation shown in the right panel of Fig. 4, where the electron can be found on the left island with probability  $p_\tau$  and on the right island with probability  $(1 - p_\tau)$ . If we measure the system at that moment, we can arrive at the level diagram shown in Fig. 5(a) (with probability  $p_\tau$ ), or at the diagram shown in Fig. 5(b) [with probability  $(1 - p_\tau)$ ]. In the first case, we quickly decrease the energy of the left island extracting the work  $V_\tau$ , and we raise the energy of the left island to the value  $V_0$  (with no cost). The average extracted work is then  $\langle W \rangle = p_\tau V_\tau$ . Then we arrive at the diagram shown in Fig. 5(c), which is a mirror of Fig. 3. It is therefore thermodynamically equivalent and we have a completed cycle. In the case shown in Fig. 5(b), we quickly move the left level up to  $V_0$  (with no cost) arriving at the situation shown in Fig. 3, again completing the cycle. There exists another protocol leading to the same consequences. Namely, at the time instance  $t = \tau$  one can quickly decrease the potential of the left island to zero (before the measurement). The extracted work is  $V_\tau$  while the probability that the left island was occupied is  $p_\tau$ , so the average work is  $\langle W \rangle = p_\tau V_\tau$ . Then we measure the position of the electron and raise the level of the empty island up to the value  $V_0$ . Again, the cycle is closed.

## IV. DERIVING THE OPTIMAL PROTOCOL

A thermodynamically equivalent turnstile version of this model was previously analyzed [23] when there were no errors in the measurements, and the consequences of reduced mutual

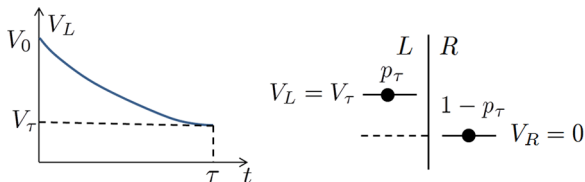


FIG. 4. Left: Example protocol during time  $0 < t < \tau$ . Right: Level diagram at  $t = \tau$ .

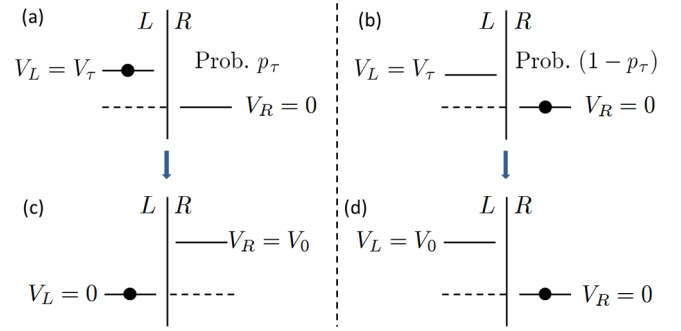


FIG. 5. Completing the cycle.

information were discussed [24]. If there is an error in the measurement, we have an additional entropy production term  $S_\varepsilon$  to the total entropy production, and the feedback operation  $V(t)$  will have to be adapted to the expected error to minimize the entropy  $\Delta S$  that is produced during feedback operation. The total entropy produced in a cycle is then given by

$$\Delta S_{\text{tot}} = \underbrace{S_\varepsilon}_{\text{measurement}} + \underbrace{\Delta S - Q/T}_{\text{operation}}, \quad (16)$$

where  $Q$  is the heat exchanged between the system and the environment. Extending the analysis from [23] to finite error is principally not difficult. We minimize the total entropy production rate,

$$\dot{S}_{\text{tot}} = \frac{\Delta S_{\text{tot}}}{\tau} = \frac{S_\varepsilon}{\tau} + \frac{\Delta H}{\tau} - P, \quad (17)$$

using the Euler-Lagrange formalism, which leads to a non-linear differential equation that has to be solved numerically. Here  $\Delta S = \Delta H$  since we assume there are no excitations from the ground state of the islands, therefore the internal entropy change is  $\Delta S_{\text{in}} = 0$ .  $P = Q/(T\tau)$  is the rate of heat exchange between the system and the environment.

### A. Thermodynamic properties

Let  $p_1(t)$  and  $p_2(t)$  be the probabilities to find the system in state 1 (the right island) and 2 (the left island), respectively. The transitions between these two states are described by the rates  $\Gamma_{12}$  and  $\Gamma_{21}$ , which satisfy the detailed balance  $\Gamma_{21}/\Gamma_{12} = e^{\Delta E/T}$ , where  $\Delta E$  is the difference in energy of the two states (note that since  $\Delta E$  is a function of time, the rates will also be time-dependent). The master equations are thus

$$\begin{aligned} \dot{p}_1 &= -\Gamma_{12}p_1 + \Gamma_{21}p_2 = -\Gamma p_1 + \Gamma_{21}, \\ \dot{p}_2 &= \Gamma_{12}p_1 - \Gamma_{21}p_2 = -\Gamma p_2 + \Gamma_{12}, \end{aligned} \quad (18)$$

where  $\Gamma(t) \equiv \Gamma_{12}(t) + \Gamma_{21}(t)$ . As in Refs. [23,25], for the sake of simplicity we choose  $\Gamma$  to be independent of time, and we set  $\Gamma = 1$ . The energy of state  $i$  is denoted  $E_i(t)$ , and in the protocol described above we have  $E_1(t) = 0$  and  $E_2(t) = V(t)$ . The total work extracted during the period  $\tau$  is

$$W_{\text{ex}} = - \sum_{i=1}^2 \int_0^\tau dt p_i \dot{E}_i, \quad (19)$$



the change in internal energy of the system is

$$\Delta U = \sum_{i=1}^2 [p_i(\tau)E_i(\tau) - p_i(0)E_i(0)], \quad (20)$$

and the transferred heat from the environment to the system is

$$Q = \Delta U + W_{\text{ex}} = \sum_{i=1}^2 \int_0^\tau dt \dot{p}_i E_i(t). \quad (21)$$

The information entropy associated with the measurement is  $H = -\sum_{i=1}^2 p_i \ln p_i$ , and the change in information entropy can be written as an integral,

$$\begin{aligned} \Delta H &= H_\tau - H_0 \\ &= -\sum_{i=1}^2 [p_i(\tau) \ln p_i(\tau) - p_i(0) \ln p_i(0)] \\ &= -\sum_{i=1}^2 \int_0^\tau dt \dot{p}_i(t) \ln p_i(t). \end{aligned} \quad (22)$$

Since  $p_1(t) = 1 - p_2(t)$ , we can relabel  $p_2(t) \equiv p$ , and we write the entropy produced per cycle as

$$\frac{\Delta H}{\tau} = -\frac{1}{\tau} \int_0^\tau dt \dot{p} \ln \left( \frac{p}{1-p} \right). \quad (23)$$

From the master equation (18), we get

$$\dot{p} = -p + \frac{1}{e^V + 1}, \quad (24)$$

where from now on we relabel  $V(t) \equiv V$ , and we measure time in units of  $\Gamma$  and energy in units of  $T$ . From this equation, we can express

$$V = \ln \left( \frac{1}{p + \dot{p}} - 1 \right).$$

The power is defined as the average heat extracted from the reservoir per cycle  $\tau$ ,  $P = Q/(T\tau)$ , and it can be written as

$$P = \frac{1}{\tau} \int_0^\tau dt \dot{p} V = \frac{1}{\tau} \int_0^\tau dt \dot{p} \ln \left( \frac{1}{p + \dot{p}} - 1 \right). \quad (25)$$

We are interested in the optimal protocol for the measurement and erasure cycle. In this system, the optimal protocol means finding the protocol  $V(t)$  and the total time  $\tau$ , which minimize the entropy production rate at a given measurement error  $\varepsilon$ . We also set the value of the power  $P$ , given by Eq. (25), to see how the solutions depend on the power we want to extract. The initial condition is set by  $p(t=0) = \varepsilon$ . That is, there is a chance,  $\varepsilon$ , that the electron was on the opposite island of what we measured.

### B. Minimizing the entropy production rate

Since we want to minimize the entropy production rate while keeping the power at a finite value  $P$ , we have to introduce the Lagrange multiplier  $\lambda$  to obtain the functional

$$J = \frac{S_\varepsilon}{\tau} + \frac{\Delta H}{\tau} + \lambda P = \frac{S_\varepsilon}{\tau} + \frac{1}{\tau} \int_0^\tau dt L(p, \dot{p}, \lambda), \quad (26)$$

with the Lagrangian

$$L(p, \dot{p}, \lambda) = \left[ -\ln \left( \frac{p}{1-p} \right) + \lambda \ln \left( \frac{1}{\dot{p} + p} - 1 \right) \right] \dot{p}. \quad (27)$$

Using the Euler-Lagrange equation

$$\frac{\partial L}{\partial p} = \frac{d}{dt} \frac{\partial L}{\partial \dot{p}}, \quad (28)$$

we obtain the following second-order nonlinear ordinary differential equation:

$$\ddot{p} = \frac{\dot{p}^2(\dot{p} + p - 1/2)}{p(\dot{p} + p - 1) + \dot{p}/2}. \quad (29)$$

To solve this equation, we need to impose a set of constraints to the solutions we want. The first constraint is that the power has to be a finite fixed value  $P$ , given by Eq. (25):

$$G(\tau, p, \dot{p}) \equiv P - \frac{1}{\tau} \int_0^\tau dt \dot{p} \ln \left( \frac{1}{p + \dot{p}} - 1 \right) = 0. \quad (30)$$

The second constraint comes from a consideration of the end-point values of  $p(t)$ . The initial condition of  $p(t)$  is given by  $p(0) = \varepsilon$ , but since the value of  $p(t)$  is not fixed at the end point  $p(\tau) = p_\tau$ , we have a second constraint,  $(\partial L / \partial \dot{p})_{t=\tau} = 0$ , which can be written as

$$\begin{aligned} F_1(\lambda, \tau, p) &= \lambda \left[ \ln \left( \frac{1}{p_\tau + \dot{p}_\tau} - 1 \right) + \frac{\dot{p}_\tau}{(\dot{p}_\tau + p_\tau - 1)(\dot{p}_\tau + p_\tau)} \right] \\ &\quad - \ln \left( \frac{p_\tau}{1 - p_\tau} \right) = 0. \end{aligned} \quad (31)$$

The third and final constraint is due to the fact that variation of Eq. (17) with respect to the period  $\tau$  should be zero. It is given by

$$\frac{\partial J}{\partial \tau} = \lambda \frac{\partial P}{\partial \tau} + \frac{\partial}{\partial \tau} \frac{\Delta H}{\tau} - \frac{S_\varepsilon}{\tau^2} = 0, \quad (32)$$

where

$$\begin{aligned} \frac{\partial P}{\partial \tau} &= \frac{\partial}{\partial \tau} \left[ \frac{1}{\tau} \int_0^\tau dt \dot{p} \ln \left( \frac{1}{p + \dot{p}} - 1 \right) \right] \\ &= \frac{\dot{p}_\tau}{\tau} \ln \left( \frac{1}{p_\tau + \dot{p}_\tau} - 1 \right) - \frac{P}{\tau} \end{aligned} \quad (33)$$

and

$$\begin{aligned} \frac{\partial}{\partial \tau} \frac{\Delta H}{\tau} &= \frac{\partial}{\partial \tau} \left[ -\frac{1}{\tau} \int_0^\tau dt \dot{p} \ln \left( \frac{p}{1-p} \right) \right] \\ &= -\frac{1}{\tau} \dot{p}_\tau \ln \left( \frac{p_\tau}{1 - p_\tau} \right) - \frac{\Delta H}{\tau^2}. \end{aligned} \quad (34)$$

The full equation for the third constraint is thus

$$\begin{aligned} F_2(\lambda, \tau, p) &\equiv \left[ \ln \left( \frac{1 - p_\tau}{p_\tau} \right) + \lambda \dot{p}_\tau \ln \left( \frac{1}{p_\tau + \dot{p}_\tau} - 1 \right) \right] \\ &\quad - \lambda P - \frac{1}{\tau} [\Delta H + S_\varepsilon] = 0. \end{aligned} \quad (35)$$

This constraint can be combined with the free-end-point constraint by eliminating the Lagrange multiplier  $\lambda$

to obtain

$$\begin{aligned}
 F(\tau, p_\tau, \dot{p}_\tau) & \\
 \equiv \ln \left( \frac{p_\tau}{1-p_\tau} \right) & [P(\dot{p}_\tau + p_\tau)(p_\tau + \dot{p}_\tau - 1) + \dot{p}_\tau^2] \\
 - \frac{S_\tau}{\tau} & \left[ \dot{p}_\tau + \ln \left( \frac{1}{p_\tau + \dot{p}_\tau} - 1 \right) \right. \\
 \times (\dot{p}_\tau + p_\tau) & \left. (p_\tau + \dot{p}_\tau - 1) \right] = 0, \quad (36)
 \end{aligned}$$

where  $S_\tau = \Delta H + S_\varepsilon = -p_\tau \ln p_\tau - (1-p_\tau) \ln(1-p_\tau)$  is the entropy of the system at time  $t = \tau$ . Euler's method is then used to solve the second-order differential equation in Eq. (29). Since it is a second-order equation, we have two constants that need to be fixed ( $\tau$  and  $V_0$ ). We find these values as the roots of the two constraints in Eqs. (30) and (36) by using Newton's method.

## V. RESULTS

We now present the main results of this analysis. The model has a parameter  $\Gamma$  that determines the tunneling rate between the two islands, and we measure time in units of  $\Gamma^{-1}$  and energy in units of temperature  $T$ . In Fig. 6 we plot the optimal period  $\tau$  as a function of the power  $P$  for selected values of the error  $\varepsilon$ .

### A. High power limit

We find that there is a maximal amount of power one can extract,  $P^{\max}(\varepsilon)$ . As this value is reached,  $\tau$  approaches zero linearly, and the entropy production rate diverges to infinity. To leading order we have

$$\tau \propto P^{\max} - P, \quad (37)$$

$$\dot{S}_{\text{tot}} \propto \frac{1}{P^{\max} - P}. \quad (38)$$

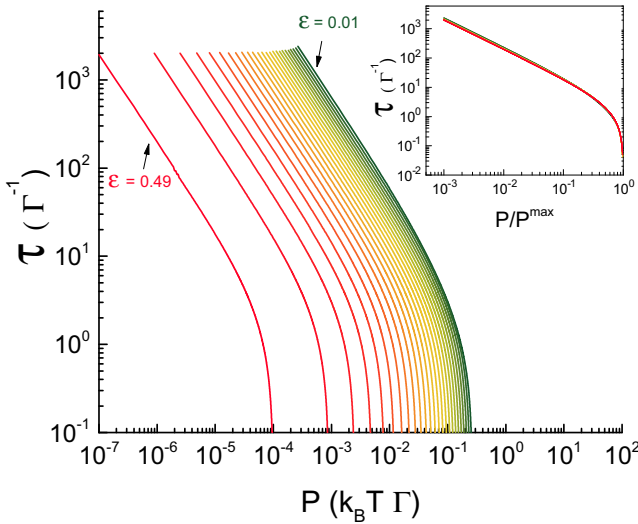


FIG. 6. The main figure shows  $\tau$  as a function of  $P$  for different  $\varepsilon$  (in steps of 0.02). The inset gives the scaled form of the same data, with  $\tau$  as a function of  $P/P^{\max}$ .

If  $\tau$  is plotted as a function of  $P/P^{\max}$ , the scaled graphs are close to collapsing over the whole range of powers, as shown in the inset of Fig. 6. We can always find the value of  $P^{\max}$  numerically, but we can also derive a single transcendental equation that determines  $P^{\max}$ , and in the case of error-free measurements we can also solve it analytically. By taking the limit as  $\tau \rightarrow 0$  in Eq. (25), we find that

$$P^{\max} = V_0 \dot{p}_0 = V_0 \left( \frac{1}{e^{V_0} + 1} - \varepsilon \right), \quad (39)$$

which expresses  $P^{\max}$  in terms of  $V_0$ . Consider Eq. (36) when  $\tau \rightarrow 0$ . Since the other terms are finite, the only way to avoid a divergence of the last term is for the expression in the large square brackets to be zero. For  $\tau = 0$  we have  $p_\tau = \varepsilon$ , and with Eq. (24) we find that  $V_0$  satisfies the equation

$$1 + (1 - V_0)e^{V_0} - \varepsilon(e^{V_0} + 1)^2 = 0. \quad (40)$$

For  $\varepsilon = 0$  we find that the maximum power is given by the Lambert  $W$  function,

$$P^{\max} = W(e^{-1}) = 0.27846 \dots \quad (41)$$

with the initial value of the potential  $V_0 = 1 + W(e^{-1})$ . This analytical result is in perfect agreement with our numerical result.

### B. Low power limit

When  $P = 0$ , we can assume that the system is always in equilibrium at the given value of  $V$ , which means that  $p = p_a = (e^V + 1)^{-1}$ . We assume for small  $P$  that we have  $p = p_a + O(P)$ , and that  $\tau = A/P$ . Inserting into Eq. (25) and expanding in  $P$ , we find that it becomes

$$1 = \frac{1}{A} \int_0^\infty dt V \dot{p}_a + O(P). \quad (42)$$

In the limit  $\tau \rightarrow \infty$ , corresponding to quasistatic operation, the entropy production will vanish if

$$(e^{V_0} + 1)^{-1} = p_0 = \varepsilon \rightarrow V_0 = \ln \left( \frac{1}{\varepsilon} - 1 \right), \quad (43)$$

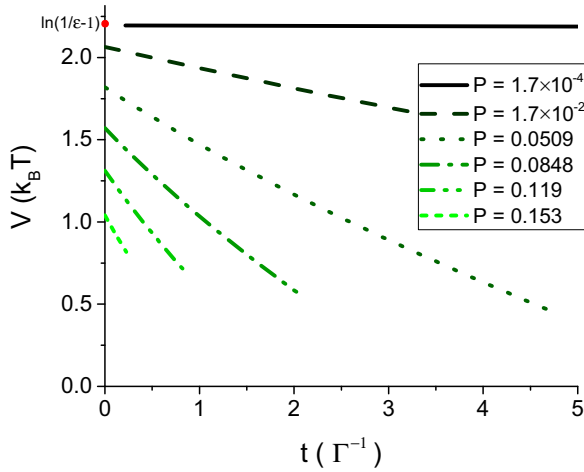
as shown in [24]. It is reasonable, and confirmed by the numerical solution, that at small  $P$  and long time  $\tau$  the potential difference will be brought all the way back to zero,  $V(\tau \rightarrow \infty) = 0$ . With these two boundary values, we get

$$A = \int_0^\infty dt V \dot{p}_a = - \int_0^{V_0} V dV \frac{dp_a}{dV} = \ln 2 - S_\varepsilon. \quad (44)$$

The end result is that in the limit of low power,  $P \rightarrow 0$ , the optimal period  $\tau$  diverges as

$$\tau = (\ln 2 - S_\varepsilon) P^{-1}, \quad (45)$$

in agreement with our numerical result. In the polynomial expansion as  $P \rightarrow 0$  of the total entropy production,  $\Delta S_{\text{tot}} = c_0 + c_1 P$ , we know that for perfect measurements  $c_0$  has to be zero since there is no entropy production during reversible operation. We have  $\tau = (\ln 2 - S_\varepsilon) P^{-1}$  for small  $P$ , and therefore we get  $\dot{S}_{\text{tot}} = \Delta S_{\text{tot}}/\tau \propto P^{-2}$ , in agreement with [23]. If errors are present, the measurement entropy  $S_\varepsilon$  exists


 FIG. 7.  $V(t)$  for  $\varepsilon = 0.1$  and several values of  $P$ .

even for reversible operation so that  $c_0 = S_\varepsilon$ , and we obtain an additional linear behavior of the entropy production rate

$$\dot{S}_{\text{tot}} = S_\varepsilon(\ln 2 - S_\varepsilon)^{-1} P + c_1^* P^2, \quad (46)$$

which we confirm numerically in Appendix 2.

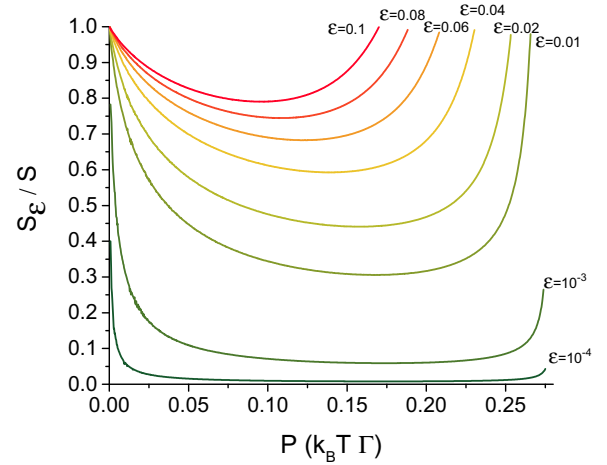
### C. The optimal protocol

Examples of optimal protocols for  $\varepsilon = 0.1$  and several values of  $P$  are shown in Fig. 7. We observe that the time  $\tau$  before the protocol should be repeated decreases with increasing  $P$ , and the initial value  $V_0$  increases with decreasing  $P$ . The quasistatic limit ( $\tau \rightarrow \infty/P \rightarrow 0$ ) found in Eq. (43) was  $V_0 = \ln(\frac{1}{\varepsilon} - 1)$ , which is marked in the plot. See Appendix 1 for more details on the behavior of  $V(t)$  and  $p(t)$  at time  $\tau$ .

To extract maximum power, one has to balance the following: the amount of energy gained per tunneling event, the probability that tunneling occurs, and the probability of back-tunneling while reducing the potential difference. These results tell us that the maximum power is reached with rapid measurements, favoring low probability high-energy tunneling events, and a steeply sloped  $V(t)$ . However, this comes at the cost of divergence in the entropy production rate. This result is obtained assuming a constant total tunneling rate  $\Gamma$ , and it may change for systems in which  $\Gamma$  depends on the potential difference between the two islands.

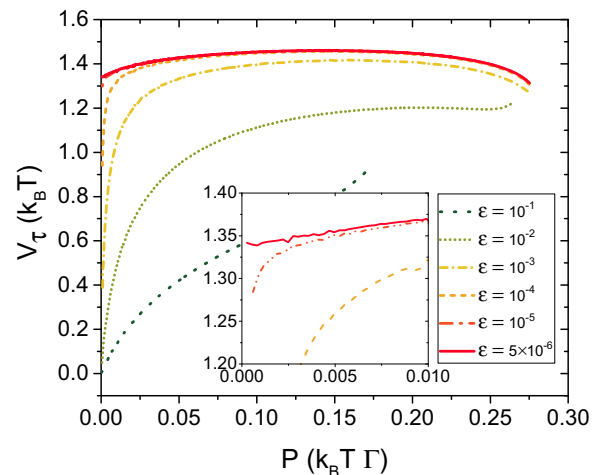
### D. Effect of the measurement error

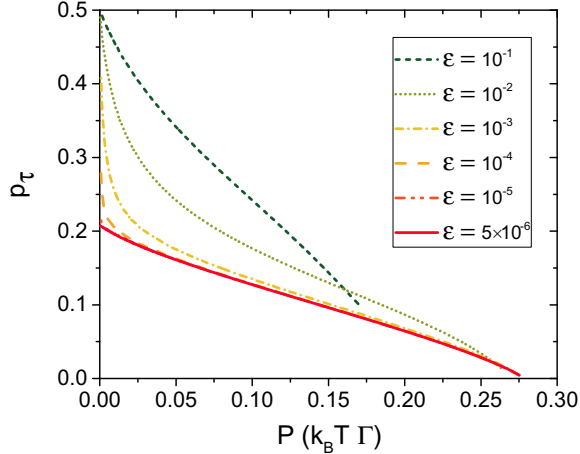
To clearly see the effect of the measurement error on the total entropy production, we plot in Fig. 8 the ratio  $S_\varepsilon/S_{\text{tot}}$  for various values of  $\varepsilon$ . For  $P \rightarrow 0$  we approach reversible operation ( $\Delta S = 0$ ), and all of the total entropy production is due to the measurement error. When  $P \rightarrow P^{\text{max}}$ , the measurement entropy dominates again since there is no time for heat transfer from the environment when  $\tau \rightarrow 0$ . For vanishingly small errors, its effect is only noticeable at the boundary values of  $P$ , but even for relatively minor measurement errors a significant portion of the entropy production is due to the measurement error for all  $P$ .


 FIG. 8. The fraction of the total entropy production  $S_{\text{tot}}$  that is due to the measurement error  $S_\varepsilon$  as a function of power for various values of  $\varepsilon$ .

## VI. SUMMARY

If we make an error in a measurement, there is an associated net entropy production. This applies to measurements of any type and with an arbitrary number of outcomes. For a symmetric binary measurement where the probability of error is  $\varepsilon$ , the entropy increases by the amount  $S_\varepsilon$ . This entropy increase can be understood from a coarse-graining of either the phase space (for a closed system) or the dynamical evolutions (for an open system). We have investigated the consequences of a finite error on the optimal performance of a realistic Szilard engine at finite (given) power. We found the existence of a maximal power  $P^{\text{max}}$ , which also exists for error-free measurements, and which decreases with increasing error. The entropy production rate diverges as the maximal power is approached. For small power, the entropy production rate is quadratic in  $P$  in the absence of errors, but it becomes linear when errors are present.


 FIG. 9.  $V_\tau$  as function of  $P$  for different  $\varepsilon$ . The inset shows enlarged what happens for small errors and powers.


 FIG. 10.  $p_\tau$  as a function of  $P$  for different  $\varepsilon$ .

### ACKNOWLEDGMENT

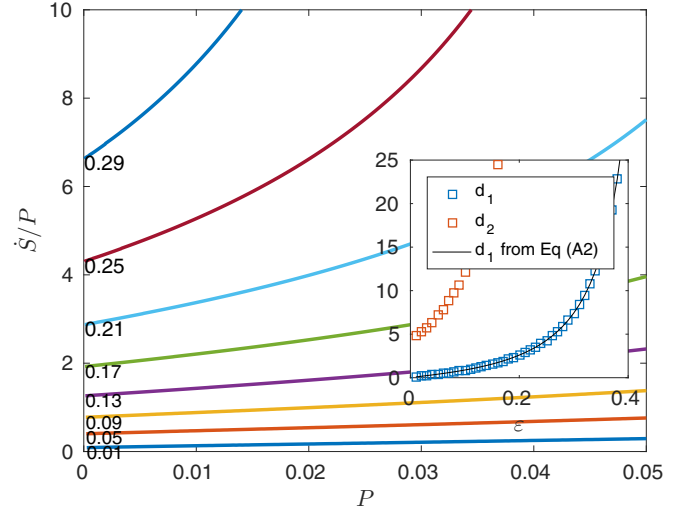
We are grateful to J. Pekola and I. Khaymovich for illuminating discussions.

### APPENDIX: ADDITIONAL RESULTS

Here we present some additional results on the optimal protocol.

#### 1. $V(t)$ and $p(t)$ at time $\tau$

Figure 9 shows  $V(\tau) = V_\tau$  as a function of  $P$  for different  $\varepsilon$ . While it seems that for any finite  $\varepsilon$  we find  $V_\tau \rightarrow 0$  as  $P \rightarrow 0$ , we see that for small  $\varepsilon$  one has to go to very small powers to see this, and for most powers  $V_\tau$  is between 1 and 1.5. This indicates a singular behavior of the function  $V_\tau(P, \varepsilon)$  at  $P = 0$  and  $\varepsilon = 0$ , and the limiting value will depend on how this point is approached. In Ref. [23] we found that  $V_\tau = 1.33$  for  $\varepsilon = 0$  and small  $P$ . From Fig. 9 (inset) we can see that this agrees well with what we would expect if we first took the limit  $\varepsilon \rightarrow 0$  and then  $P \rightarrow 0$ . The same singularity is reflected in the probability  $p_\tau$  to find the electron on the opposite island at time  $\tau$  from the one it was measured at time 0 as shown


 FIG. 11.  $\dot{S}/P$  as a function of  $P$  with labels on the curves giving  $\varepsilon$ . For each curve, the value at  $P = 0$  and the slope of the tangent at that point will give the coefficients  $c_1$  and  $c_2$  of Eq. (A1). These are shown as functions of  $\varepsilon$  in the inset, together with  $c_1$  from Eq. (A2).

in Fig. 10. For all finite  $\varepsilon$  we have  $\lim_{P \rightarrow 0} p_\tau = 0.5$ , but for small  $\varepsilon$  this only happens at very small  $P$ .

#### 2. Polynomial expansion of $\dot{S}$

In Ref. [23] it was found that for  $\varepsilon = 0$  and small  $P$ ,  $\dot{S}$  is proportional to  $P^2$ . We find that this is not true for finite  $\varepsilon$ . We expand to second order,

$$\dot{S} = d_1 P + d_2 P^2, \quad (\text{A1})$$

where  $d_1$  and  $d_2$  are functions of  $\varepsilon$ . Plotting  $\dot{S}/P$  as a function of  $P$  (Fig. 11), we get  $d_1$  and  $d_2$  as the intercept and slope of the tangent at  $P \rightarrow 0$  (Fig. 11, inset). From the analytical results in Eq. (46), we know that  $d_1 = S_\varepsilon (\ln 2 - S_\varepsilon)^{-1}$ ,

$$\dot{S} = d_1 P \quad \text{with} \quad d_1 = S_\varepsilon (\ln 2 - S_\varepsilon)^{-1}, \quad (\text{A2})$$

which, as shown in Fig. 11 (inset), agrees perfectly with the numerical solution.

- 
- [1] *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*, edited by H. S. Leff and A. Rex (Institute of Physics, Bristol, 2003).
- [2] L. Szilard, Über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen, *Z. Phys.* **53**, 840 (1929).
- [3] L. Brillouin, *Science and Information Theory* (Academic Press, New York, 1956).
- [4] R. Landauer, Irreversibility and heat generation in the computing process, *IBM J. Res. Dev.* **5**, 183 (1961).
- [5] C. H. Bennett, The thermodynamics of computation—A review, *Int. J. Theor. Phys.* **21**, 905 (1982).
- [6] J. D. Norton, Waiting for Landauer, *Stud. Hist. Philos. Sci. Pt. B* **42**, 184 (2011).
- [7] J. Ladyman and K. Robertson, Landauer defended: Reply to Norton, *Stud. Hist. Philos. Sci. Pt. B* **44**, 263 (2013).
- [8] T. Sagawa and M. Ueda, Minimal Energy Cost for Thermodynamic Information Processing: Measurement and Information Erasure, *Phys. Rev. Lett.* **102**, 250602 (2009).
- [9] G. N. Price, S. T. Bannerman, K. Viering, E. Narevicius, and M. G. Raizen, Single-Photon Atomic Cooling, *Phys. Rev. Lett.* **100**, 093004 (2008).
- [10] 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).
- [11] M. G. Raizen, Comprehensive control of atomic motion, *Science* **324**, 1403 (2009).
- [12] 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).



- [13] A. Berut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, Experimental verification of Landauer's principle linking information and thermodynamics, *Nature (London)* **483**, 187 (2012).
- [14] V. Serreli, C.-F. Lee, E. R. Kay, and D. A. Leigh, A molecular information ratchet, *Nature (London)* **445**, 523 (2007).
- [15] J. V. Koski, V. F. Maisi, J. P. Pekola, and D. V. Averin, Experimental realization of a Szilard engine with a single electron, *Proc. Natl. Acad. Sci. USA* **111**, 13786 (2014).
- [16] 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).
- [17] K. Chida, K. Nishiguchi, G. Yamahata, H. Tanaka, and A. Fujiwara, Thermal-noise suppression in nano-scale Si field-effect transistors by feedback control based on single-electron detection, *Appl. Phys. Lett.* **107**, 073110 (2015).
- [18] 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).
- [19] T. Sagawa and M. Ueda, Role of mutual information in entropy production under information exchanges, *New J. Phys.* **15**, 125012 (2013).
- [20] K. Ridderbos, The coarse-graining approach to statistical mechanics: How blissful is our ignorance? *Stud. Hist. Philos. Sci. Pt. B* **33**, 65 (2002).
- [21] J. M. Blatt, An alternative approach to the ergodic problem, *Prog. Theor. Phys.* **22**, 745 (1959).
- [22] S. Lloyd, Use of mutual information to decrease entropy: Implications for the second law of thermodynamics, *Phys. Rev. A* **39**, 5378 (1989).
- [23] J. Bergli, Y. M. Galperin, and N. B. Kopnin, Information flow and optimal protocol for a Maxwell-demon single-electron pump, *Phys. Rev. E* **88**, 062139 (2013).
- [24] J. M. Horowitz and J. M. R. Parrondo, Thermodynamic reversibility in feedback processes, *Europhys. Lett.* **95**, 10005 (2011).
- [25] D. Mandal and C. Jarzynski, Work and information processing in a solvable model of Maxwell's demon, *Proc. Natl. Acad. Sci. USA* **109**, 11641 (2012).