PHYSICS

New fluctuation theorems on Maxwell's demon

Qian Zeng[1] **and Jin Wang**[2]*

With increasing interest in the control of systems at the nano- and mesoscopic scales, studies have been focused on the limit of the energy dissipation in an open system by refining the concept of the Maxwell's demon. To uncover the underlying physical principle behind a system controlled by a demon, we prove a previously unexplored set of fluctuation theorems. These fluctuation theorems imply that there exists an intrinsic nonequilibrium state of the system, led by the nonnegative demon-induced dissipative information. A consequence of this analysis is that the bounds of both work and heat are tighter than the limits predicted by the Sagawa-Ueda theorem. We also suggest a possible experimental test of these work and heat bounds.

INTRODUCTION

In the history of physics, the well-known Maxwell's demon was proposed to act as a rebel against the authority of the thermodynamic second law (*1*). It decreases the entropy in a thermally isolated system and finally rescues the whole universe from the heat death. Despite the myth of its existence (*2*, *3*), the demon reflects the habitus of the universal system, especially at microscales: The system interacting with a demon becomes open and thus behaves far away from thermal equilibrium. There is a deep connection between the nonequilibrium thermodynamics involved in the Maxwell's demon and the information theory, such as in the much-studied cases of Szilard engine (*4*) and Landauer principle (*5*, *6*). The physical nature of information may be revealed by the study on the demon. For this reason, many efforts have been devoted to this direction. The related works have shown their importance in the theoretic and experimental areas of nano- and mesoscopic system analysis and control (*7*–*15*).

As a central concept in modern thermodynamics, the entropy production quantifies the energy dissipation in a stochastic system. One of the fundamental properties of the entropy production is that it follows the Jarzynski equality (*16*) or the integral fluctuation theorem, which is regarded as the generalized second law from a microscopic perspective. To analyze the demon's effect, several pioneering works attempted to construct an improper entropy production, which disobeys the Jarzynski equality (*17*–*22*). This thought follows the original idea of Maxwell. One representative of this construction was given by Sagawa and Ueda (*23*, *24*), where a fluctuation theorem (Sagawa-Ueda theorem) has been developed for the improper entropy production by taking into account the information acquired by the demon. Correspondingly, a generalized second law arises from this fluctuation theorem: The demon cannot extract work more than the acquired information on average. This result gives plausible interpretation on the Szilard's engine and many other models, respectively. However, there are still unsolved problems in the frameworks of this kind for the following reasons.

First, the improper entropy production arises because the system dynamics is measured in an inconsistent manner where a part of the demon's contribution is missing. Thus, the improper entropy production measures the energy dissipation incorrectly. Intuitively, the demon controls not only the system state but also the energy exchanges, such as the work and heat between the system and the baths. Thus, the demon contributes to the entropies in both the system and the baths. With this thought, one can construct different improper entropy productions by neglecting any part of the demon's contribution (from either the system or the baths, or parts of them). Correspondingly, there exist different fluctuation theorems for these entropy productions, which can lead to different second law inequalities for work or heat. The first question is which inequality is more appropriate? Second, the equality in a second law inequality always represents the thermal equilibrium state of the system. However, a system is supposed to be in a nonequilibrium state when controlled by a demon. This indicates that if the demon works efficiently, then the equality in the second law in previous frameworks does not always hold. It has been reported by several works (*25*–*28*) in the examples of the information processing that the upper bound of the extracted work is less than the bound predicted by the Sagawa-Ueda theorem. This reveals the fact that when the system is at a controlled nonequilibrium state, there exists an additional energy dissipation, which is not estimated by the previous frameworks. The second question is where this energy dissipation originated?

The motivation of this paper is to draw a clearer picture of the Maxwell's demon. We note the fact that the controlled system actually follows the second law when the dynamics is properly measured. One can quantify the correct entropy productions at different coarse-grained levels for the demon's control. Every improper entropy production can give rise to a missing part of the demon's contribution. None of these entropy productions fulfills the task of complete characterization of the demon unless we take the total contribution into account. The puzzle of the demon obviously involves the interactions between the system and the demon during the whole dynamics. In the thermodynamics, it is appropriate to describe these interactions by using the informational correlation—the dynamical mutual information (*29*–*31*) defined as $i = \log\frac{p[x(t)\mid y(t)]}{p[x(t)]}$, where $x(t)$ and $y(t)$ represent the two simultaneous trajectories of the two interacting systems, respectively, and p denotes the probability (density) of the trajectories. With this quantification at the trajectory level, it is natural to introduce the concept of dissipative information (*32*–*36*) to quantify the time irreversibility of the dynamical mutual information

$$\sigma_I = i - \tilde{i} \quad (1)$$

[1]State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Changchun, Jilin 130022, China. [2]Departments of Chemistry and of Physics and Astronomy, State University of New York, Stony Brook, NY 11794-3400, USA.
*Corresponding author. Email: jin.wang.1@stonybrook.edu

where $\tilde{i} = \log \frac{p[\tilde{x}(t)|\tilde{y}(t)]}{p[\tilde{x}(t)]}$ is the dynamical mutual information along the time-reversed trajectories. We will show that σ_I rightly quantifies the demon's total contribution. For a complete thermodynamical description, one should develop a set of fluctuation theorems, which includes not only the entropy production in the system but also the dissipative information, rather than the construction of improper entropy productions. The fluctuation theorems on the entropy productions reflect the nonequilibrium dynamics in the controlled system. Different from the ordinary fluctuation theorems for one single system, the fluctuation theorem on the dissipative information quantifies the nonequilibrium interactions or binary relations. It is thus reasonable to believe that when the demon works efficiently, there exists an intrinsic nonequilibrium state (due to the binary relations) characterized by a positive averaged dissipative information. This is the source of the inevitable energy dissipation in many cases of the demon.

RESULTS

Fluctuation theorems and inequalities

Let us consider that a demon controls a system that is coupled with several thermal baths. The system and the demon are initially at the states x_0 and y, respectively. Then, the demon performs a control to the system with a protocol $\Gamma(y)$ based on y. For simplicity, the correspondence between y and $\Gamma(y)$ is assumed to be bijective. Consequentially, the system's trajectory $x(t)$ is correlated to the demon state y. As a reasonable assumption, the demon does not alter the control protocol while the demon state y is unchanged during the dynamics. Driven by thermal baths, the stochasticity of the system allows the time-reversal trajectory $\tilde{x}(t) \equiv x(\tau - t)$ to be under the identical protocol. Here, the initial state of $\tilde{x}(t)$ corresponds exactly to the final state of $x(t)$ denoted by x_t.

When $\Gamma(y)$ or y is displayed explicitly in the system dynamics, an entropy production can be given by log ratio between the probabilities (densities) of $x(t)$ and $\tilde{x}(t)$ conditioning on y

$$\sigma_{X|Y} = \log \frac{p[x(t)|y]}{p[\tilde{x}(t)|y]} = \Delta s_{X|Y} + \delta s_{X|Y} \tag{2}$$

where the subscript $X|Y$ means that the thermodynamical entity of the system (X) is controlled by a given protocol of the demon (Y). In addition, $\sigma_{X|Y}$ can be viewed as the total stochastic entropy change consisting of the contributions from the system and the baths at the microscopic level (*37*). This is because the total entropy change can be given by the second equality in Eq. 2. Here, $\Delta s_{X|Y} = -\log p(x_t|y) - [-\log p(x_0|y)]$ quantifies the stochastic entropy difference of the system between the final and initial states; $\delta s_{X|Y} = \log \frac{p[x(t)|x_0, y]}{p[\tilde{x}(t)|x_t, y]}$ represents the stochastic entropy flow from the system to the baths, which is also identified as the heat transferred from the baths to the system as $Q_{X|Y} = -T\delta s_{X|Y}$, which has been proven in the detailed fluctuation theorem in the Langevin or Markovian dynamics (*38*, *39*). Thus, $\delta s_{X|Y}$ is recognized as the (stochastic) entropy change in the baths. On the other hand, when the demon's control $\Gamma(y)$ or the demon state y is unknown in the system dynamics, the entropy production can be measured properly at the coarse-grained level. That is to say, one needs to average or integrate the demon's control information out of the dynamics, i.e., to obtain the marginal probability $p[x(t)] = \sum_y p(y) p[x(t)|y]$ with implicit control conditions. Then, another entropy production, which is a coarse-grained version of $\sigma_{X|Y}$, can be given by

$$\sigma_X = \log \frac{p[x(t)]}{p[\tilde{x}(t)]} = \Delta s_X + \delta s_X \tag{3}$$

In the second equality in Eq. 3, $\Delta s_x = \log \frac{p(x_0)}{p(x_t)}$ and $\delta s_X = \log \frac{p[x(t)|x_0]}{p[\tilde{x}(t)|x_t]}$ are recognized as the coarse-grained entropy changes in the system and in the baths, respectively. Thus, σ_X quantifies the total entropy change at the coarse-grained level with the lack of the demon's control information. An illustrative case for showing the differences between the entropy productions can be found in Fig. 1. It is interesting that both $\sigma_{X|Y}$ and σ_X follow the Jarzynski equalities

$$\langle \exp(-\sigma_{X|Y}) \rangle = 1, \text{and} \langle \exp(-\sigma_X) \rangle = 1 \tag{4}$$

where the average $\langle \exp(-\sigma_{X|Y}) \rangle$ is taken over the ensembles of the system and the demon's state. One should note that for every protocol, $\sigma_{X|Y}$ obeys the detailed Jarzynski equality under every possible control protocol, i.e., $\langle \exp(-\sigma_{X|Y}) \rangle_{X|Y} = 1$, where the average $\langle \cdot \rangle_{X|Y}$ is taken over the ensemble of the system while y is fixed. For a complete view of the controlled nonequilibrium thermodynamics of the system, it is appropriate to take the average of the detailed Jarzynski equality on both sides over the ensemble of the demon's state with the notation $\langle \cdot \rangle \equiv \langle \langle \cdot \rangle_{X|Y} \rangle_Y$. Notice that together, the two Jarzynski equalities in Eq. 4 provide a new sight that the second law holds for the system at both two levels of the knowledge of demon's control.

In general, the two entropy productions shown above are different from each other. The gap between them indicates the demon's contribution to entropy production, which is exactly the dissipative information σ_I shown in Eq. 1, where the trajectory $y(t)$ is fixed

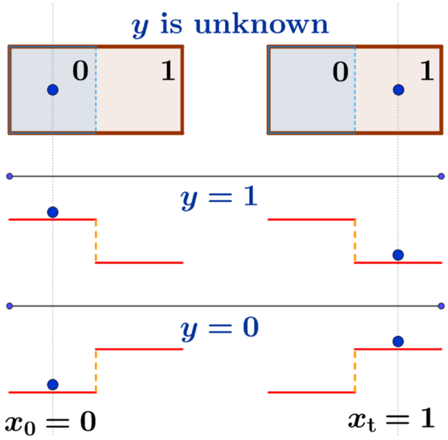

Fig. 1. The entropy productions at the finest and coarse-grained levels under the demon's control. A particle (shown as the blue circle) is confined in a box. The state of the particle can be represented by 0 or 1 when the particle is contained in the corresponding half of the box. A demon controls the particle system by exerting different potentials to the system. A trajectory of the particle in the position representation is given by $x(t) = \{x_0 = 0, x_t = 1\}$. In the first row, the detailed information of the potential is unknown and the entropy production σ_X can only be measured by using the coarse-grained dynamics. In the second row, the demon exerts an explicit potential to the system corresponding to $y = 1$, and the entropy production at the fine level is given by $\sigma_{X|y=1}$ at the finest level. In the third row, the demon exerts another potential explicitly, and the entropy production is given by $\sigma_{X|y=0}$. The three entropy productions are not equal to each other in general.

at a single value of state y. This can be seen from the following relationship

$$\sigma_{X\mid Y} = \sigma_X + \sigma_I \tag{5A}$$

The detailed contributions of the demon to the system and the baths can be revealed by the decomposition of dissipative information and the relations between the entropy changes shown in Eqs. 2 and 3 in the following equalities

$$\begin{cases} \sigma_I = \Delta i + \delta i \\ \Delta s_{X\mid Y} = \Delta s_X + \Delta i \\ \delta s_{X\mid Y} = \delta s_X + \delta i \end{cases} \tag{5B}$$

Here, $\Delta i = i_0 - i_t$ is the information change of the system during the dynamics, with $i_0 = \log \frac{p(x_0\mid y)}{p(x_0)}$ and $i_t = \log \frac{p(x_t\mid y)}{p(x_t)}$ being the state mutual information between the system state and the demon's state at initial and final time, respectively, which has been introduced in (*40*, *41*); $\delta i = \rho - \tilde{\rho}$ is the time-irreversible information transfer from the demon to the system. Here, $\rho = \log \frac{p[x(t)\mid x_0, y]}{p[x(t)\mid x_0]}$ and $\tilde{\rho} = \log \frac{p[\tilde{x}(t)\mid x_t, y]}{p[\tilde{x}(t)\mid x_t]}$ quantify the information transferred (*42*–*44*) from the demon to the system along the forward-in-time and backward-in-time trajectories, respectively. The information transfer is an informational measure of how the dynamics of the system depends on the demon by using the comparison between the system dynamics at different coarse-grained levels under the demon's control ($p[x(t)\mid x_0, y]$ and $p[x(t)\mid x_0]$). In Eq. 5B, the second equality identifies the role of Δi that it can be regarded as the demon's contribution to the entropy change in the system; the third equality indicates that δi depicts the demon's contribution to the baths. Then, the role of dissipative information is clear: It describes how the demon influences the entropy production through the nonequilibrium binary relation or interaction (see Fig. 2). Moreover, this effect can be quantified precisely in the following fluctuation theorem

$$\langle \exp(-\sigma_I) \rangle = 16 \tag{6}$$

This is a new fluctuation theorem, which is quite different from the Jarzynski equality because it is for the nonequilibrium thermodynamics of the binary interactions between the systems rather than for a single system.

To resolve the puzzle of the demon, we first review the construction of the improper entropy productions. A construction, $\eta = \Delta s_X + \delta s_{X\mid Y}$ is an improper entropy production because it violates

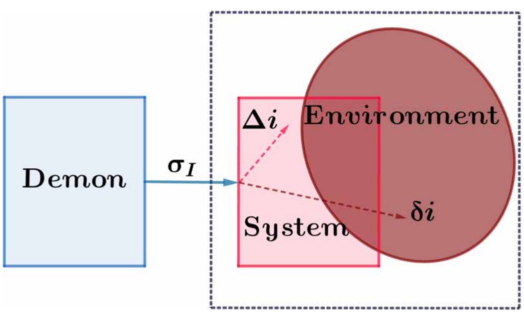

Fig. 2. Detailed contributions of the demon to the entropy changes in the system (denoted by Δi) and the baths (denoted by δi), respectively.

Jarzynski equality $\langle \exp(-\eta)\rangle \neq 1$, and the two entropy changes Δs_X and $\delta s_{X\mid Y}$ are measured at different levels of the knowledge of the demon's control according to Eqs. 2 and 3. On the other hand, η arises because Δi is neglected in $\sigma_{X\mid Y}$, $\eta = \sigma_{X\mid Y} - \Delta i$, suggested by Eq. 5B. This indicates that in Eq. 4, the Jarzynski equality for $\sigma_{X\mid Y}$ can be satisfied by adding the contribution of Δi to η. This is the core of Sagawa-Ueda theorem, which emphasizes $\Delta I = -\Delta i$ as the key characterization of the demon. Following the similar idea, one can construct different improper entropy productions. For instance, consider $\eta' = \Delta s_{X\mid Y} + \delta s_X$, where $\Delta s_{X\mid Y}$ and δs_X are measured in an inconsistent manner in the dynamics, thus $\langle \exp(-\eta')\rangle \neq 1$. By adding δi into η', one has $\sigma_{X\mid Y} = \eta' + \delta i$, which gives rise to the same Jarzynski equality for $\sigma_{X\mid Y}$ in Eq. 4. However, neither Δi nor δi quantifies the total demon's contribution because Δi gives the demon's influence on the system, while δi gives the demon's influence on the baths. Therefore, only the dissipative information σ_I involving both the demon's control on the system and baths can take into account the overall contribution of the demon. Unlike in previous works, the relation in Eq. 5A, together with the corresponding set of fluctuation theorems in Eqs. 4 and 6, provides the full clear picture of Maxwell's demon.

We further derive a series of inequalities to obtain the bounds on the dissipative entities (entropy productions and dissipative information). By applying Jensen's inequality $\langle \exp(-O)\rangle \geq \exp(-\langle O\rangle)$ to Eqs. 4 and 6, respectively, we have

$$\begin{cases} \langle \sigma_{X\mid Y}\rangle \geq 0, \text{or } \langle \Delta s_{X\mid Y}\rangle \geq -\langle \delta s_{X\mid Y}\rangle \\ \langle \sigma_X\rangle \geq 0, \text{or} \langle \Delta s_X\rangle \geq -\langle \delta s_X\rangle \\ \langle \sigma_I\rangle \geq 0, \text{or } \langle \Delta i\rangle \geq -\langle \delta i\rangle \end{cases} \tag{7}$$

The first two are the second law inequalities at different coarse-grained levels of the demon's control corresponding to the Jarzynski equalities in Eq. 4, while the last inequalities about σ_I shows the previously unknown feature of the nonequilibrium behavior brought by the demon. To see this, take the average on both sides of Eq. 5A over the ensembles; we have $\langle \sigma_{X\mid Y}\rangle = \langle \sigma_X\rangle + \langle \sigma_I\rangle$. Combining with Eq. 7, one sees that $\langle \sigma_{X\mid Y}\rangle$ quantifies the true (utmost) entropy productions in the system. A lower bound of $\langle \sigma_{X\mid Y}\rangle$ different from that obtained from the second law in Eq. 7 (which is zero) is given by the following inequality

$$\langle \sigma_{X\mid Y}\rangle \geq \langle \sigma_I\rangle \geq 0 \tag{8}$$

In Eq. 8, $\langle \sigma_{X\mid Y}\rangle = 0$ at the finest level indicates that the system is in a quasi-static (equilibrium) process, where every control protocol is applied infinitely slowly. Such a demon does not work efficiently in practice. High efficiency means achieving the control in a finite time, which leads to a nonequilibrium process. Consequentially, the lower bound of $\langle \sigma_{X\mid Y}\rangle$ is always a positive number rather than 0. Although measured properly, $\langle \sigma_X\rangle$ does not reflect the true nonequilibrium thermodynamics of the system because of the coarse-graining. Meanwhile, $\langle \sigma_X\rangle$ does not need to be strictly positive when the system is actually in nonequilibrium. However, there always exists a positive dissipative information ($\langle \sigma_I\rangle > 0$), which is contained in the true entropy production $\langle \sigma_{X\mid Y}\rangle$. This is due to the nonequilibrium part of the dynamical mutual information for the binary relationship between the demon and the system. The exceptions can be seen in the cases where a demon controls the system with a unique and deterministic protocol or noise-free protocols; we have $\langle \sigma_I\rangle = 0$ as

$\langle\sigma_X\rangle = \langle\sigma_{X\mid Y}\rangle$ during the dynamics. Otherwise, there exists an intrinsic nonequilibrium state of the system in general, which is characterized by an inevitable energy dissipation given by $\langle\sigma_{X\mid Y}\rangle = \langle\sigma_I\rangle > 0$.

We should stress that the dissipative information given by Eq. 1 may fail to work as a dissipative entity when the demon state can change during the dynamics. This is because there is no fluctuation theorem or inequality that can guarantee the universal nonnegativity of the average of Eq. 1 (*32*, *34*). However, we can show that with the proper categorization of the individual subdynamics of the system and the demon from the total dynamics, the fluctuation theorems in Eqs. 4 and 6 and the inequalities in Eqs. 7 and 8 still hold for the system and the demon in the subdynamics, respectively. This indicates the universality of the presented fluctuation theorems and the inequalities about the demon model. The subdynamics categorization method and the detailed derivations are shown in section S3.

New bounds for work and heat

A consequence of Eq. 8 is that the bounds on the heat and work should be revised beyond the second law. To see this, let us assume that the system is coupled with a thermal bath with temperature T for simplicity. Then, the system dynamics can be given by the Langevin dynamics. The Hamiltonian of the system depends on the system state and the control protocol denoted by $H(x, y) \equiv H(x, \Gamma(y))$ [y and $\Gamma(y)$ is one-to-one correspondence to each other]. The change in the Hamiltonian, depending on concrete demon state, can be given by $\Delta H_{X\mid Y} = H(x_t, y) - H(x_0, y)$. With the assumption of the local detailed balance condition (*38*), the entropy production can be given in terms of the stochastic heat absorbed by the system $\sigma_{X\mid Y} = \Delta s_{X\mid Y} - T^{-1}Q_{X\mid Y}$, where the heat can be given by $Q_{X\mid Y} = -T\delta s_{X\mid Y}$. According to the thermodynamic first law, stated as $\Delta H_{X\mid Y} = Q_{X\mid Y} + W_{X\mid Y}$ with $W_{X\mid Y}$ being the stochastic work performed on the system, $\sigma_{X\mid Y}$ can be rewritten in terms of the work $\sigma_{X\mid Y} = T^{-1}[W_{X\mid Y} - \Delta F_{X\mid y}]$. Here, $\Delta F_{X\mid y}$ is the Helmholtz free energy difference depending on the concrete demon state y, given by $\Delta F_{X\mid y} = \langle\Delta H_{X\mid Y}\rangle_{X\mid Y} - T\langle\Delta s_{X\mid Y}\rangle_{X\mid Y}$. The probability weights in the averages of the state variables in $\Delta F_{X\mid y}$ should be distinguished at the initial and final states: The weights $p(x_0\mid y)$ and $p(x_t\mid y)$ with concrete demon state y are used for x_0 and x_t, respectively. Then, according to the inequality for $\sigma_{X\mid Y}$ in Eq. 7, we reach the ordinary second law inequalities for the heat and work

$$\langle Q_{X\mid Y}\rangle \leq T\langle\Delta s_{X\mid Y}\rangle, \text{and } \langle W_{X\mid Y}\rangle \geq \Delta F_{X\mid Y} \tag{9}$$

Here, $\Delta F_{X\mid Y} = \langle\Delta F_{X\mid y}\rangle_Y = \langle\Delta H_{X\mid Y}\rangle - T\,\langle\Delta s_{X\mid Y}\rangle$ is recognized as the averaged free energy difference of the system controlled by the demon, and $\Delta F_{X\mid Y}$ no longer depends on concrete demon states.

Note that different constructions of improper entropy productions can lead to the same form of the second law in Eq. 9. For example, the Sagawa-Ueda theorem suggested to use the coarse-grained entropy change $\langle\Delta s_X\rangle$ instead of the true entropy change $\langle\Delta s_{X\mid Y}\rangle$ but still use the true heat and work given in Eq. 9 in the statement of the second law. The reason is that the entropy change $\langle\Delta s_X\rangle$ and the true heat and work in Eq. 9 can be measured properly in practice. On the other hand, the correlation between the demon and the system may be unknown in practice. This correlation can be quantified by the information change $\langle\Delta i\rangle = \langle\Delta s_{X\mid Y}\rangle - \langle\Delta s_X\rangle$ by noting the relation in Eq. 5B. Then, the averaged heat $\langle Q_{X\mid Y}\rangle$ in Eq. 9 can sometimes be greater than the coarse-grained entropy change, i.e., $\langle Q_{X\mid Y}\rangle \geq T\langle\Delta s_X\rangle$, and the second law is seemingly violated. Following the Sagawa-Ueda theorem, the free energy difference can be given as follows (*40*)

$$\Delta F = \langle\Delta H_{X\mid Y}\rangle - T\langle\Delta s_X\rangle = \Delta F_{X\mid Y} + T\langle\Delta i\rangle$$

where the averaged change in the Hamiltonian $\langle\Delta H_{X\mid Y}\rangle$ is used the same as in Eq. 9 but with a different entropy change $\langle\Delta s_X\rangle$. Then, the averaged work $\langle W_{X\mid Y}\rangle$ in Eq. 9 does not satisfy the second law seemingly, since $\langle W_{X\mid Y}\rangle$ can sometimes be less than the free energy difference ΔF, i.e., $\langle W_{X\mid Y}\rangle \leq \Delta F$. According to the Sagawa-Ueda theorem, by adding the information change into the entropy change $\langle\Delta s_X\rangle$ and the free energy difference ΔF, the averaged heat and work can satisfy the following second law inequalities

$$\begin{aligned}\langle Q_{X\mid Y}\rangle &\leq T\langle\Delta s_X\rangle + T\langle\Delta i\rangle = T\langle\Delta s_{X\mid Y}\rangle\\ \langle W_{X\mid Y}\rangle &\geq \Delta F - T\langle\Delta i\rangle = \Delta F_{X\mid Y}\end{aligned}$$

This can clarify the equivalence between the second law inequalities given in Eq. 9 and those given by the Sagawa-Ueda theorem. Furthermore, the terms $\langle\Delta s_{X\mid Y}\rangle$ and $\Delta F_{X\mid Y}$ in Eq. 9 can be recognized as the true/complete entropy change and the true/complete free energy difference with the complete knowledge of the demon control. In this sense, we refer to Eq. 9 as the complete form of the second law. On the other hand, all the improper entropy productions are generated by decomposing the total entropy production $\sigma_{X\mid Y}$ in different ways, and thus, they can lead to the same complete form in Eq. 9.

Now, we take the dissipative information into account. By noting Eq. 8, we reach tighter bounds for the heat and the work compared to Eq. 9

$$\begin{aligned}\langle Q_{X\mid Y}\rangle &\leq T\langle\Delta s_{X\mid Y}\rangle - T\langle\sigma_I\rangle \leq T\langle\Delta s_{X\mid Y}\rangle = T\langle\Delta s_X\rangle + T\langle\Delta i\rangle\\ \langle W_{X\mid Y}\rangle &\geq \Delta F_{X\mid Y} + T\langle\sigma_I\rangle \geq \Delta F_{X\mid Y} = \Delta F - T\langle\Delta i\rangle\end{aligned} \tag{10}$$

where in the looser bounds (the last equalities in Eq. 10), the free energy difference ΔF, the information change $\langle\Delta i\rangle$, and the entropy change $\langle\Delta s_X\rangle$ are used in the Sagawa-Ueda theorem, as shown above. Here, we obtain a smaller upper bound for the heat and a larger lower bound for the work than the ordinary second law in Eq. 9, which can also be formulated by the Sagawa-Ueda theorem. These tighter bounds clearly indicate the nontrivial nonequilibrium state of a system controlled by a demon. On the basis of the above discussion, the looser bounds of the heat and work in the second inequalities (see also Eq. 9) represent the equilibrium limit while the demon does not work efficiently. However, when the environments are complex and noisy, the interacting systems have to pay positive amounts of energy dissipation costs to maintain the connections to each other. These necessary energy dissipations can show how far the systems are away from the equilibrium but cannot be predicted by (the complete form of) the second law in Eq. 9. This is because the second law (and other generalized forms) is based on the entropy production, which can only provide the equilibrium bound (looser bounds in Eq. 10) for the total dissipation. On the other hand, the dissipative information can quantify the energy dissipations led by the interactions by the amount of $T\langle\sigma_I\rangle$. This yields the tighter bounds in Eq. 10. Here, we stress again the looser bounds in Eq. 10 (the last equalities) can be predicted by the Sagawa-Ueda theorem by taking the information change into account, as discussed above.

Usually in the practical model of Maxwell's demon such as the Szilard's type demon, the action of the demon is divided into two different processes: measurement and feedback control. In the measurement process, the demon observes the system and acquires the information of the system state. The demon needs to match its state with the system during the measurement, and the system can be viewed as the outer controller of the demon. Here, we can use X to present the demon and Y to denote the system (observed by the demon) in the measurement process. In this situation, an inevitable heat from the demon X to its environmental bath, or say, the measurement heat Q_{mea} can be generated from the demon X during the information acquirement (*45*, *46*), which is the negative heat $(-Q_{X\mid Y})$ from the bath to the demon X, $Q_{mea} = -Q_{X\mid Y} = T\delta s_{X\mid Y}$, where $\delta s_{X\mid Y}$ is the entropy change in the bath. In the feedback control process (we let demon $=Y$ and controlled system $=X$), the demon Y extracts a positive work W_{ext} from the system X with an additional energy dissipation, which is the negative work performed on the system $W_{ext} = -W_{X\mid Y}$. Then, the bounds for Q_{mea} and W_{ext} can be given by the complete form of the second law in Eq. 9, where the equalities hold for infinitely slow quasi-static or equilibrium processes. However, if the demon works efficiently, then we come to a nonequilibrium situation where an amount of positive energy dissipation is necessary and originated from the dissipative information. Thus, new bounds for the heat Q_{mea} from the demon and the work W_{ext} from the system can be given by Eq. 10 in the form of

$$\begin{aligned}\langle Q_{\mathrm{mea}}\rangle &\geq T\langle\sigma_I\rangle - T\langle\Delta s_{X\mid Y}\rangle \geq -T\langle\Delta s_{X\mid Y}\rangle = -T\langle\Delta s_X\rangle - T\langle\Delta i\rangle \\ \langle W_{\mathrm{ext}}\rangle &\leq -\Delta F_{X\mid Y} - T\langle\sigma_I\rangle \leq -\Delta F_{X\mid Y} = -\Delta F + T\langle\Delta i\rangle\end{aligned} \tag{11}$$

In Eq. 11, the equivalent relationship between the Sagawa-Ueda theorem and the complete form of the second law has been shown in the last equalities, where ΔF and $\langle\Delta s_X\rangle$ denote the free energy difference and the entropy change with incomplete knowledge of the demon. We give the interpretation about Eq. 11 as follows.

The information change $\langle\Delta i\rangle = I_0 - I_t$ can often be negative in the measurement process because the final mutual information between the system and the demon is always larger than the initial mutual information. If the initial mutual information I_0 and the entropy change $\langle\Delta s_X\rangle$ can be neglected in the measurement process, then we have that the true entropy change is equal to the information obtained from the system, $\langle\Delta s_{X\mid Y}\rangle = -I_t$. For example, the demon and the system are often uncorrelated at the beginning of the measurement process, where $I_0 = 0$. In addition, the demon can be in the same equilibrium state at the initial and final time of the coarse-grained dynamics, and then we have $\langle\Delta s_X\rangle$ (see the measurement case in the "Illustrative Cases" section below). Then, the inequality for the measurement heat reduces to $\langle Q_{mea}\rangle \geq T\langle\sigma_I\rangle + TI_t \geq TI_t$ in Eq. 11. The looser bound TI_t is the minimal energy requirement for the demon to obtain the information I_t from the system in the equilibrium or noise-free (no-measurement-error) limit. The dissipative information $T\langle\sigma_I\rangle$ quantifies the additional energy cost for the demon to establish the correlation to the system under the nonequilibrium and noisy condition. Thus, we have the tighter lower bound $\langle\sigma_I\rangle + TI_t$ for the heat in Eq. 11. This means that there is more heat generated in the measurement than the estimations given by the complete form of the second law.

On the other hand, the information change $\langle\Delta i\rangle = I_0 - I_t$ can usually be positive in the feedback control process. This is because the demon can use the (initially) obtained information I_0 to extract work, and the correlation between the demon and the system can decrease after the control $I_t < I_0$. If the final mutual information I_t and the incomplete free energy difference ΔF can vanish in the control process, then we find that the initial information I_0 works as the true/complete free energy difference in this situation, $\Delta F_{X\mid Y} = -TI_0$. For example, in the model of the Szilard's type demon, the controlled system is assumed to be at the same equilibrium state at the initial and final time of the dynamics, and then we have the incomplete free energy difference $\Delta F = 0$. This equilibrium state of the system can be independent of the demon. The demon and the system are correlated right after the initial time $I_0 > 0$ and finally uncorrelated at the end, where $I_t = 0$. Then, we have the inequality for the work from Eq. 11 as $\langle W_{\mathrm{ext}}\rangle \leq TI_0 - T\langle\sigma_I\rangle \leq TI_0$. The looser bound for the work, TI_0, has been predicted by the Sagawa-Ueda theorem, which represents that the work can be extracted under the equilibrium or noise-free condition. However, the tighter upper bound $TI_0 - T\langle\sigma_I\rangle$ indicates that there should be a waste of information that cannot be used to extract work in the nonequilibrium and noisy operations. This wasted information can be quantified by the dissipative information $\langle\sigma_I\rangle$. This means that there is less work extracted in the feedback control than the estimations given by the Sagawa-Ueda theorem.

Illustrative cases

To illustrate our idea in this paper, we calculate the cases of the information ratchets shown in Fig. 3, which can be tested in the experiments. A potential with the two wells is exerted on a confined

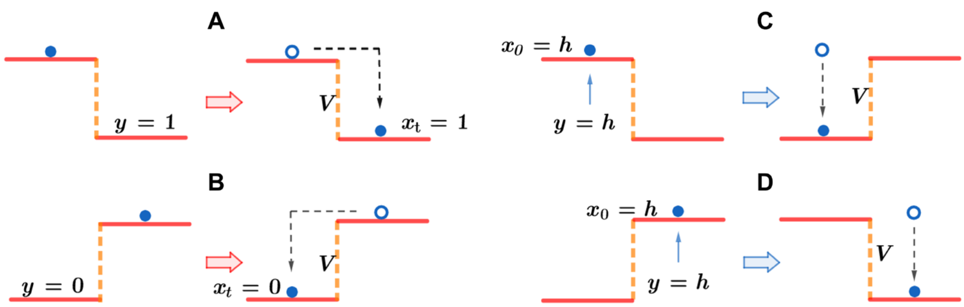

Fig. 3. The confined particle works as a demon or is controlled by a demon. In (**A**) and (**B**), the particle is used as a demon and measures the system state before the control. The system state, denoted by y, is represented by the location of the lower well (0 or 1) in the potential. In (A), $y = 1$; and in (B), $y = 0$. The final state of the particle, denoted by x_t, is taken as the state of y. In (**C**) and (**D**), the particle is controlled by a demon. The initial state of the particle, denoted by x_0, is represented by h or l when the particle is at the higher or the lower well. When spotting the state $x_0 = h$, the demon reverses the potential and extracts a positive work of V from the particle system.

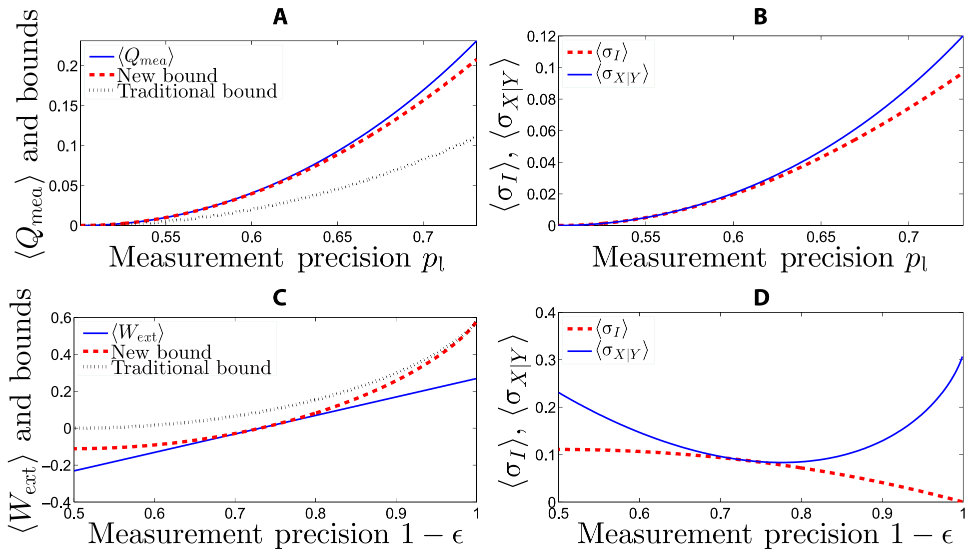

Fig. 4. The bounds for the measurement heat (Eq. 12), the extracted work (Eq. 13), and the true entropy productions (Eq. 8) of the illustrative cases. The traditional bounds are given by the Sagawa-Ueda theorem, and the tighter bounds are provided by the presented fluctuation theorems. Due to the Sagawa-Ueda theorem, the traditional lower bounds for the true entropy productions are given by zeros. In (**A**), the averaged measurement heat $\langle Q_{mea}\rangle$ (solid line), the traditional lower bound I_t (dotted line), and the new lower bound $I_t + \langle \sigma_I \rangle$ (dash line) in the measurement are plotted as functions of the measurement precision p_l. The potential height V is raised from 0 to 1. Correspondingly, p_l is increased from 0.50 to 0.73 monotonically. The corresponding dissipative information $\langle \sigma_I \rangle$ (dash line) and the entropy production $\langle \sigma_{X|Y} \rangle$ (solid line) in the measurement are shown as the functions of the precision p_l in (**B**). In (**C**), the extracted work $\langle W_{mea}\rangle$ (solid line), the traditional upper bound I_0 (dotted line), and the new upper bound $I_0 - I_c$ (dash line) are plotted as functions of the measurement precision $1-\epsilon$, where $1-\epsilon$ is ranged from 0.5 to 1, and the potential height is $V = 1$. The corresponding dissipative information $\langle \sigma_I \rangle$ (dash line) and the entropy production $\langle \sigma_{X|Y} \rangle$ (solid line) are shown in (**D**).

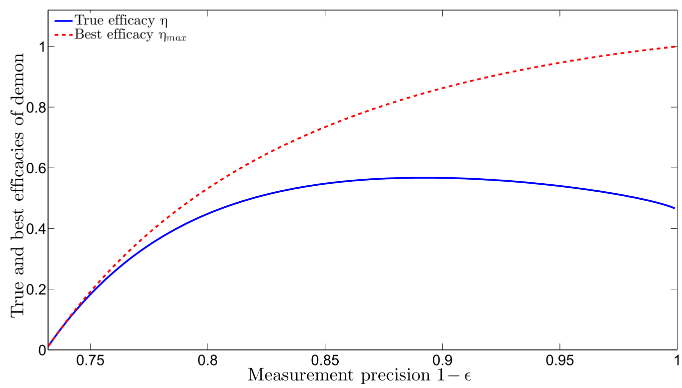

Fig. 5. The true efficacy $\eta = \frac{\langle W_{ext}\rangle}{I_0}$ and the best efficacy $\eta_{max} = 1 - \frac{I_c}{I_0}$ of the demon given in Eq. 15A and Eq. 15B. The measurement precision is characterized by the probability $p(y = x_0 \mid x_0) = 1 - \epsilon$. The extracted work W_{ext} can be positive and monotonically increasing in the range of the measurement precision $p(y = x_0 \mid x_0) \in [0.73,1]$ in this case. Meanwhile, the best efficacy η_{max} of the demon also monotonically increases (goes up to 1) in this range of the measurement precision, and η_{max} works as the upper bound of the true efficacy η.

particle. The height between the two wells is equal to $V > 0$. While under the equilibrium, the probabilities that the particle is at the lower and the higher well can be quantified by $p_l = [1 + \exp(-V)]^{-1}$ and $p_h = 1 - p_l$, respectively, ($p_l > 1/2$). An outside controller can control the particle by reversing the profile of the potential, i.e., by raising the lower well up to V and lowering the higher well down to 0. The action of the controller is assumed to be fast enough before the particle reacts. For no loss of generality, the temperature of the environmental bath is assumed to be at $T = 1$.

If the particle works as a demon [shown in Fig. 3 (A and B)], then the particle is supposed to measure the state of the controlled system at first. The state of the particle is denoted by $x = 0$ or 1 when at the left or the right well, respectively. The system state can be represented by the location of the lower well with the value of $y = 0$ or 1 with equal probability $p(y = 0) = p(y = 1) = 1/2$. The particle is initially under the equilibrium until the system state changes. Correspondingly, the potential is reversed by the system immediately, and the particle starts measuring the current system state. When the equilibrium is achieved, the final state x_t of the particle is taken as an observation of y. The probability of the measurement error can be given by the probability of the particle at the higher well $p_{Xt\mid Y}(x_t \neq y \mid y) = p_h$. On the other hand, the measurement precision is characterized by the probability of the particle at the lower well $p_{Xt\mid Y}(x_t = y \mid y) = p_l$. By noting the definitions and relationships shown in Eq. 2, the averaged measurement heat generated by the particle can be given by $\langle Q_{mea}\rangle = (1/2 - p_h)V$, and the entropy change can be evaluated by $\langle \Delta s_{X\mid Y}\rangle = -I_t$ (see eqs. S26 to S28). Here, $I_t = \log 2 - S \geq 0$ is the final mutual information that measures the correlation between the observation x_t and the state y (*40*, *41*), where the Shannon entropy S is given by $S = -p_l \log p_l - p_h \log p_h$. Then, according to Eq. 11 and as shown in the above discussions, the tighter bound of $\langle Q_{mea}\rangle$ in this case can be given by

$$\langle Q_{\rm mea}\rangle \geq \langle \sigma_I \rangle + I_t \geq I_t \quad (12)$$

Here, the dissipative information $\langle \sigma_I \rangle$ can be calculated by using the probabilities of the forward and backward trajectories $x(t) = \{x_0, x_t\}$ and $\tilde{x}(t) = \{x_t, x_0\}$, respectively. By inserting these probabilities into Eq. 1, we have the expression $\langle \sigma_I \rangle = \log \sqrt{2p_l^2 + 2p_h^2} \geq 0$ (see eq. S29). Although the measurement precision characterized by p_l

increases as the potential height V increases, higher precision also raises up both the averaged measurement heat and the lower bound of the energy dissipation quantified by the dissipative information in this case. The numerical results can be found in Fig. 4 (A and B).

Next, we use a demon to extract positive work from the particle system [shown in Fig. 3 (C and D)]. In this case, the state of the particle can be denoted by $x = l$ or h when at the lower or the higher well, respectively. Initially, the particle is under equilibrium. The demon measures the state of the particle at first and obtains the observation y. The demon plays the feedback control according to the observation y. When the particle is observed to be at the higher well, the demon reverses the potential immediately and extracts an amount of work $W_{ext} = V$. After the control, the demon does nothing until the particle goes to the equilibrium again. For a practical thought, the demon's measurement can have a random error, and this error certainly lowers the efficiency of the work extraction. Here, we simply assume that the measurement error occurs with stable probability $p_{Y\mid X0}(y \neq x_0 \mid x_0) = \epsilon$. By using Eq. 2 and noting the thermodynamical first law, the extracted work can be given by $\langle W_{ext}\rangle = (p_h - \epsilon)V$, on average, and the true/complete free energy difference is equal to the mutual information change during the dynamics, $\Delta F_{X\mid Y} = -I_0$ (see eqs. S30 to S33). Here, the incomplete free energy difference suggested by the Sagawa-Ueda theorem (see Eq. 11) vanishes in this case, $\Delta F = 0$. The mutual information $I_0 = S_Y - S\epsilon \geq 0$ represents the initial correlation between the demon and the particle, where the Shannon entropies can be given by $S_Y = -p_y \log p_y - (1 - p_y)\log(1 - p_y)$ and $S\epsilon = -\epsilon\log\epsilon - (1 - \epsilon)\log(1 - \epsilon)$, with $p_y = p_l(1 - \epsilon) + p_h\epsilon$ representing the probability of the observation $y = l$. Then, because of Eq. 11 and as shown in the above discussions, the bound of $\langle W_{ext}\rangle$ can be given by

$$\langle W_{\text{ext}}\rangle \leq I_0 - I_c \leq I_0 \tag{13}$$

Here, the mutual information $I_c = S_Y - S \geq 0$ measures correlation between the demon and the particle right after the control, where the Shannon entropy is $S = -p_l\log p_l - p_h\log p_h$. It is important to note that I_c is actually the information that is not used to extract the work but can merely dissipate into the bath as the dissipative information $\langle\sigma_I\rangle = I_c$. This result can be verified by evaluating Eq. 1 (see eq. S18). For this reason, the demon can only extract work less than the mutual information difference before and after the control, quantified by $I_0 - I_c$. We can see that higher measurement precision characterized by $1 - \epsilon$ can increase the averaged extracted work (with fixed potential height V); meanwhile, the inevitable dissipative information is decreased by the increasing precision in this case. The numerical results are shown in Fig. 2 (C and D). Also, we can note that the dissipative information $\langle\sigma_I\rangle$ bounds the entropy production $\langle\sigma_{X\mid Y}\rangle$ from the below in both the cases of the measurement and feedback control [see Fig. 4 (B and D)]. This verifies the inequality in Eq. 8.

On the other hand, we find that the tighter upper bound in Eq. 13 is equivalent to the information process second law (*26*–*28*) by noting $I_0 - I_c = S - S\epsilon$. Here, S and $S\epsilon$ can be regarded as the Shannon entropies of a "0,1" tape before and after the information processing, respectively (*25*). This indicates that the proposed fluctuation theorems in this paper can be applied to the area of thermodynamics computing from a general perspective. In addition, the looser bounds for the heat and work in Eqs. 12 and 13 are predicted by the second law (Sagawa-Ueda theorem), and these bounds can only be achieved in the quasi-static (or equilibrium) or noise-free control protocols.

Here, we also relate our case of the work extraction to the efficacy parameter γ, which appeared in the experimental and theoretical works (*8*, *23*). Here, γ can be shown as the departure of the fluctuation of an improper entropy production $\sigma = \sigma_{X\mid Y} - i_0$ from unity, i.e.,

$$\langle\exp(-\sigma)\rangle = \gamma \neq 1 \tag{14A}$$

where $\sigma_{X\mid Y}$ is the total (proper) entropy production and satisfies the fluctuation theorem in Eq. 4, and i_0 is the initial stochastic mutual information with average $\langle i_0\rangle = I_0 \geq 0$. The final mutual information I_t is supposed to vanish after the control. The parameter γ was interpreted as the efficacy parameter of the demon, and the demon can behave better under a high value of γ. However, that appears to be not the case. By applying the Jensen inequality to Eq. 14A, we can obtain the following inequality for γ,

$$\langle\sigma_{X\mid Y}\rangle \geq I_0 - \log\gamma \tag{14B}$$

where the identity $\langle\sigma_{X\mid Y}\rangle = \langle\sigma\rangle - I_0$ has been used in Eq. 14B. We see that the lower bound in Eq. 14B can be negative, especially at the high values of γ. However, a negative bound of the total (or proper) entropy production $\langle\sigma_{X\mid Y}\rangle$ cannot be achieved. Thus, the parameter γ appears to be lack of the physical meaning of the efficacy of the demon in this situation. More detailed discussions and derivations can be found in the "Meaning and bounds of the parameter γ" section in section S2. From the perspective of our theory and by noting Eq. 13, the following ratio can be used to quantify the demon's best efficacy that can be achieved instead of γ

$$\eta_{\max} = 1 - \frac{I_c}{I_0} \tag{15A}$$

From Eq. 13, $\eta_{\max}$ satisfies the inequality as follows

$$\eta = \frac{\langle W_{\text{ext}}\rangle}{I_0} \leq \eta_{\max} \leq 1 \tag{15B}$$

where η is the ratio of the extracted work $\langle W_{ext}\rangle$ to the maximum work I_0 that can be extracted in the quasi-static or noise-free limit. Thus, η measures the true efficacy of the demon in the work extraction. Equation 15B gives a clear physical meaning to the ratio η_{max}:η_{max} quantifies ratio of the maximum work $I_0 - I_c$ to the quasi-static or noise-free limit work I_0, with the minimal wasted (or dissipative) information (I_c) considered. In the presented case of the work extraction, $\eta_{\max} = 1 - \frac{I_c}{I_0} = \frac{S - S_\epsilon}{S_Y - S_\epsilon}$ can be shown as a monotonically increasing function of the precision probability of the measurement $p_{Y\mid X0}(y = x_0 \mid x_0) = 1 - \epsilon$. Here, $y = x_0$ indicates the correct measurement. The meanings of the entropies S, S_Y, and $S\epsilon$ have been provided in the above case of the work extraction. When the error-free condition $p(y = x_0 \mid x_0) = 1$ is achieved, the best efficacy η_{max} can go up to 1 (see Fig. 5). Thus, this η_{max} follows the intuition that the measurement accuracy can enhance the efficacy of the work extraction.

DISCUSSION

Traditional analysis on the Maxwell's demon focuses on how the second law is violated by the system and is rescued by some hidden demon-induced entities. These entities were believed as the key characterization of the demon. In contrast, we show that the system does not disobey the second law whether the demon is hidden or not, which can be seen in the set of fluctuation theorems (Eq. 4) for the entropy productions when they are correctly measured (Eqs. 2 and 3). Intrinsically, the nonequilibrium behavior of the system led by the demon is due to the time-irreversibility of the binary relationship between them, which is quantified by the dissipative information (Eq. 1). In addition, we prove another new fluctuation theorem for this dissipative information (Eq. 6). This theorem (Eq. 6) combining with the other fluctuation theorems (Eq. 4) for the entropy productions gives a precise quantification of the effect of the demon. An apparent result following these theorems is that there exists an inevitable energy dissipation originated from the positive dissipative information, which leads to the tighter bounds for the work and the heat (Eq. 11) than that estimated by the complete form of the second law or the Sagawa-Ueda theorem (Eq. 9). We also suggest a possible realization of the experimental estimation of these work and heat bounds, which can be measured and tested. These results offer a general picture of a large class of the models of the Maxwell's demon.

MATERIALS AND METHODS

Proof of the fluctuation theorems

The probabilities (densities) $p[x(t)\,|\,y]$ and $p[x(t)]$ are assumed to be nonnegative and to be normalized, i.e., $p[x(t)\,|\,y]$, $p[x(t)] \geq 0$, respectively, $\int p[x(t)\,|\,y]Dx(t) = 1$ and $\int p[x(t)]Dx(t) = 1$. Besides, we need that the differentials, with respect to the time-forward and backward trajectories, are equal to each other, i.e., $Dx(t) = D\tilde{x}(t)$. For the entropy productions and the dissipative information in Eqs. 1 to 3, we obtain the equalities

$$\langle \exp(-\sigma_{X\,|\,Y}) \rangle = \int dy \int p(y)\, p[x(t)|y] \frac{p[\tilde{x}(t)|y]}{p[x(t)|y]} Dx(t) = 1$$

$$\langle \exp(-\sigma_X) \rangle = \int p[x(t)] \frac{p[\tilde{x}(t)]}{p[x(t)]} Dx(t) = 1$$

$$\langle \exp(-\sigma_I) \rangle = \int dy \int p[x(t)]\, p[y\,|\,x(t)] \frac{p[y\,|\,\tilde{x}(t)]}{p[y\,|\,x(t)]} Dx(t) = 1$$

In the last equation for σ_I, by noting the relation in the probabilities that $p[y\,|\,x(t)] = \frac{p(y)\,p[x(t)\,|\,y]}{p[x(t)]}$, we have $\sigma_I = i - \tilde{i} = \log \frac{p[\tilde{y}\,|\,x(t)]}{p[y\,|\,\tilde{x}(t)]}$. This completes the proof on the fluctuation theorems in Eqs. 4 and 6.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/23/eabf1807/DC1

REFERENCES AND NOTES

1. J. C. Maxwell, *Theory of Heat* (Greenwood Press, ed. 3, 1970).
2. O. R. Shenker, Maxwell's Demon 2: Entropy, classical and quantum information computing. *Stud. Hist. Philos. Sci. B* **35**, 537–540 (2004).
3. K. Maruyama, F. Nori, V. Vedral, Colloquium: The physics of Maxwell's demon and information. *Rev. Mod. Phys.* **81**, 1–23 (2009).
4. L. Szilard, über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter Wesen. *Z. Physik* **53**, 840–856 (1929).
5. R. Landauer, Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **5**, 183–191 (1961).
6. M. Esposito, C. Van den Broeck, Second law and Landauer principle far from equilibrium. *Europhys. Lett.* **95**, 40004 (2011).
7. S. Deffner, C. Jarzynski, Information processing and the second law of thermodynamics: An inclusive, Hamiltonian approach. *Phys. Rev. X* **3**, 041003 (2013).
8. S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, M. Sano, Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nat. Phys.* **6**, 988–992 (2010).
9. D. Mandal, C. Jarzynski, Work and information processing in a solvable model of Maxwell's demon. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11641–11645 (2012).
10. J. V. Koski, A. Kutvonen, I. M. Khaymovich, T. Ala-Nissila, J. P. Pekola, On-chip Maxwell's demon as an information-powered refrigerator. *Phys. Rev. Lett.* **115**, 260602 (2015).
11. V. Serreli, C. F. Lee, E. R. Kay, D. A. Leigh, A molecular information ratchet. *Nature* **445**, 523–527 (2007).
12. M. Alvarez-Perez, S. M. Goldup, D. A. Leigh, A. M. Z. Slawin, A chemically driven molecular information ratchet. *J. Am. Chem. Soc.* **130**, 1836–1838 (2008).
13. A. B. Boyd, J. P. Crutchfield, Maxwell demon dynamics: Deterministic chaos, the Szilard map, and the intelligence of thermodynamic systems. *Phys. Rev. Lett.* **116**, 190601 (2016).
14. T. McGrath, N. S. Jones, P. R. ten Wolde, T. E. Ouldridge, Biochemical machines for the interconversion of mutual information and work. *Phys. Rev. Lett.* **118**, 028101 (2017).
15. P. P. Potts, P. Samuelsson, Thermodynamic uncertainty relations including measurement and feedback. *Phys. Rev. E* **100**, 052137 (2019).
16. C. Jarzynski, Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **78**, 2690–2693 (1997).
17. M. Esposito, G. Schaller, Stochastic thermodynamics for "Maxwell demon" feedbacks. *Europhys. lett.* **99**, 30003 (2012).
18. P. Strasberg, G. Schaller, T. Brandes, M. Esposito, Quantum and information thermodynamics: A unifying framework based on repeated interactions. *Phys. Rev. X* **7**, 021003 (2017).
19. J. M. Horowitz, M. Esposito, Thermodynamics with continuous information flow. *Phys. Rev. X* **4**, 031015 (2014).
20. K. Ptaszynski, M. Esposito, Thermodynamics of quantum information flows. *Phys. Rev. Lett.* **122**, 150603 (2019).
21. J. M. Horowitz, H. Sandberg, Second-law-like inequalities with information and their interpretations. *New J. Phys.* **16**, 125007 (2014).
22. D. Hartich, A. C. Barato, U. Seifert, Stochastic thermodynamics of bipartite systems: Transfer entropy inequalities and a Maxwell's demon interpretation. *J. Stat. Mech. Theory Exp.* **2014**, P02016 (2014).
23. T. Sagawa, M. Ueda, Generalized Jarzynski equality under nonequilibrium feedback control. *Phys. Rev. Lett.* **104**, 090602 (2010).
24. J. M. R. Parrondo, J. M. Horowitz, T. Sagawa, Thermodynamics of information. *Nat. Phys.* **11**, 131–139 (2015).
25. A. C. Barato, U. Seifert, Unifying three perspectives on information processing in stochastic thermodynamics. *Phys. Rev. Lett.* **112**, 090601 (2014).
26. A. B. Boyd, D. Mandal, J. P. Crutchfield, Identifying functional thermodynamics in autonomous Maxwellian ratchets. *New J. Phys.* **18**, 023049 (2016).
27. C. Aghamohammadi, J. P. Crutchfield, Thermodynamics of random number generation. *Phys. Rev. E* **95**, 062139 (2017).
28. A. B. Boyd, M. Dibyendu, J. P. Crutchfield, Thermodynamics of modularity: Structural costs beyond the Landauer bound. *Phys. Rev. X* **8**, 031036 (2018).
29. T. M. Cover, J. A. Thomas, *Elements of Information Theory* (Wiley, ed. 2, 2006).
30. F. Tostevin, P. Rein ten Wolde, Mutual information between input and output trajectories of biochemical networks. *Phys. Rev. Lett.* **102**, 218101 (2009).
31. A. C. Barato, D. Hartich, U. Seifert, Rate of mutual information between coarse-grained non-Markovian variables. *J. Stat. Phys.* **153**, 460–478 (2013).
32. G. Diana, M. Esposito, Mutual entropy production in bipartite systems. *J. Stat. Mech. Theory Exp.* **16**, P04010 (2014).
33. Q. Zeng, J. Wang, Information landscape and flux, mutual information rate decomposition, and connections to entropy production. *Entropy* **19**, 678 (2017).
34. Q. Zeng, J. Wang, Non-Markovian nonequilibrium information dynamics. *Phys. Rev. E* **98**, 032123 (2018).
35. X. Fang, K. Kruse, T. Lu, J. Wang, Nonequilibrium physics in biology. *Rev. Mod. Phys.* **91**, 045004 (2019).
36. P. P. Potts, P. Samuelsson, Detailed fluctuation relation for arbitrary measurement and feedback schemes. *Phys. Rev. Lett.* **121**, 210603 (2018).
37. U. Seifert, Stochastic thermodynamics, fluctuation theorems, and molecular machines. *Rep. Prog. Phys.* **75**, 126001 (2012).
38. G. E. Crooks, Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E* **60**, 2721–2726 (1999).
39. C. Jarzynski, Hamiltonian derivation of a detailed fluctuation theorem. *J. Stat. Phys.* **98**, 77–102 (2000).
40. T. Sagawa, M. Ueda, Fluctuation theorem with information exchange: Role of correlations in stochastic thermodynamics. *Phys. Rev. Lett.* **109**, 180602 (2012).
41. T. Sagawa, M. Ueda, Role of mutual information in entropy production under information exchange. *New J. Phys.* **15**, 125012 (2013).
42. T. Schreiber, Measuring information transfer. *Phys. Rev. Lett.* **85**, 461–464 (2000).

43. S. Ito, Backward transfer entropy: Informational measure for detecting hidden Markov models and its interpretations in thermodynamics, gambling and causality. *Sci. Rep.* **6**, 36831 (2016).
44. R. E. Spinney, J. T. Lizier, M. Prokopenko, Transfer entropy in physical systems and the arrow of time. *Phys. Rev. E* **94**, 022135 (2016).
45. C. H. Bennett, The thermodynamics of computation – a review. *Int. J. Theor. Phys.* **21**, 905–940 (1982).
46. T. Sagawa, M. Ueda, Minimal energy cost for thermodynamic information processing: measurement and information erasure. *Phys. Rev. Lett.* **102**, 250602 (2009).
47. K. Sekimoto, *Stochastic Energetics* (Springer, 2010).

Acknowledgments
Funding: Q.Z. thanks the support of National Natural Science Foundation of China (NSFC-91430217). **Author contributions:** Q.Z. carried out the mathematical derivations, performed the analysis, and wrote the manuscript. J.W. initiated and supervised the research, performed the analysis, and wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 9 October 2020
Accepted 20 April 2021
Published 4 June 2021
10.1126/sciadv.abf1807

Citation: Q. Zeng, J. Wang, New fluctuation theorems on Maxwell's demon. *Sci. Adv.* **7**, eabf1807 (2021).