Nonequilibrium thermodynamics of feedback control

Takahiro Sagawa[1,2] and Masahito Ueda[3,4]

[1]*Hakubi Center, Kyoto University, Yoshida-ushinomiya cho, Sakyo-ku, Kyoto 606-8302, Japan*

[2]*Yukawa Institute for Theoretical Physics, Kyoto University, Kitashirakawa-oiwake cho, Sakyo-ku, Kyoto 606-8502, Japan*

[3]*Department of Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

[4]*ERATO Macroscopic Quantum Control Project, JST, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

(Received 16 May 2011; revised manuscript received 13 December 2011; published 3 February 2012)

We establish a general theory of feedback control on classical stochastic thermodynamic systems and generalize nonequilibrium equalities such as the fluctuation theorem and the Jarzynski equality in the presence of feedback control with multiple measurements. Our results are generalizations of the previous relevant works to the situations with general measurements and multiple heat baths. The obtained equalities involve additional terms that characterize the information obtained by measurements or the efficacy of feedback control. A generalized Szilard engine and a feedback-controlled ratchet are shown to satisfy the derived equalities.

DOI: 10.1103/PhysRevE.85.021104 PACS number(s): 05.70.Ln, 82.60.Qr, 05.20.−y

I. INTRODUCTION

Since the mid-twentieth century, feedback control has played crucial roles in science and engineering [1,2]. Here "feedback" means that a control protocol depends on measurement outcomes obtained from the controlled system. Recently feedback control has become increasingly important in terms of nonequilibrium physics, due to at least the following two reasons.

First, stochastic aspects of thermodynamics [3–6] have become important due to recent theoretical and experimental developments. Theoretically, a number of nonequilibrium equalities such as the fluctuation theorem and the Jarzynski equality [7–41] have recently been found. On the other hand, experimental techniques have been developed to manipulate and observe small thermodynamic systems such as macromolecules and colloidal particles, and several nonequilibrium equalities have been experimentally verified [42–52]. Moreover, artificial [53–56] and biological [57] molecular machines have been investigated. In these contexts, feedback control is useful to realize intended dynamical properties of small thermodynamic systems, and it has become a topic of active research [58–82].

Second, feedback control sheds light on the foundations of thermodynamics and statistical mechanics concerning "Maxwell's demon" [83–88]. In fact, Maxwell's demon performs measurement and feedback control on thermodynamic systems. Recently Maxwell's demon has attracted renewed interest [89–104] from the standpoints of modern information theory and statistical mechanics.

A quintessential model of Maxwell's demon is a single-particle heat engine proposed by Szilard in 1929 [85]. During the thermodynamic cycle of the Szilard engine, the demon obtains 1 bit ($=\ln 2$ nat) of information by a measurement, performs feedback control, and extracts $k_{\rm B}T\ln 2$ of positive work from a single heat bath. After numerous arguments on the consistency between the demon and the second law of thermodynamics, it is now understood that the work needed for the demon (or equivalently the feedback controller) during the measurement and information erasure compensates for the work that can be extracted by the demon [102]. Therefore, we cannot extract a net positive work from the total system of the engine and the demon in an isothermal cycle, and therefore the presence of the demon does not contradict the second law of thermodynamics. Nevertheless, $k_{\rm B}T\ln 2$ of work extracted by the demon can still be useful. By using feedback control, we can increase the system's free energy without injecting any energy (work) to it. We stress that, without feedback control, we need the direct energy input into the system in order to increase its free energy due to the second law of thermodynamics. Feedback control may be regarded as a powerful tool to control thermodynamic systems. Since the crucial quantity is the information that is obtained to be used for feedback control, we may regard the Szilard-type heat engine as an "information heat engine." Recently such an information heat engine was realized experimentally by using a colloidal particle [82].

In this paper we formulate a general theory of feedback control on stochastic thermodynamic systems. In particular, we extend recent theoretical results on the generalizations of the fluctuation theorem and the Jarzynski equality [67] to the situations in which the measurement and feedback control are non-Markovian and there are multiple heat baths. Our results serve as the fundamental building blocks of information heat engines.

This paper is organized as follows.

In Sec. II we briefly review the framework of stochastic thermodynamics in a general setup. We discuss classical stochastic systems that are in general non-Markovian and in contact with multiple heat baths. We discuss the concept of entropy production and the detailed fluctuation theorem as our starting point. Because they are general properties of nonequilibrium systems, our formulations and results in the following sections are not restricted to Langevin systems but applicable to any classical stochastic systems that satisfy the detailed fluctuation theorem.

In Sec. III we formulate measurements on thermodynamic systems. We discuss multiple measurements, including continuous measurements, and investigate the properties of the mutual information obtained by the measurements. In particular, we introduce the mutual information (or the transfer

entropy) I_c, which will be shown to play key roles in the discussion of feedback control.

In Sec. IV we discuss feedback control on Markov and non-Markov processes and investigate feedback control in terms of probability theory, where the causality of the measurement and feedback play a crucial role.

In Sec. V we derive the main results of this paper. We generalize the nonequilibrium equalities to situations in which the system is subject to feedback control. In particular, we derive two types of generalizations of the fluctuation theorem and the Jarzynski equality. One involves a term concerning the mutual information, and the other involves a term of feedback efficacy. As corollaries, we derive the generalizations of the second law of thermodynamics and a fluctuation-dissipation relation.

In Sec. VI we illustrate our general results by two examples: a generalized Szilard engine with measurement errors and a feedback-controlled ratchet [58,61,63]. We discuss the former analytically and the latter numerically.

In Sec. VII we conclude this paper.

In the Appendix, we discuss the physical meaning of entropy production to elucidate the physical contents of our results in two typical situations.

II. REVIEW OF STOCHASTIC THERMODYNAMICS

In this section we briefly review thermodynamics of classical stochastic systems and introduce notations that will be used later.

A. Dynamics

We consider a classical stochastic system $\mathbf{S}$ that is in contact with heat baths $\mathbf{B}_1$, $\mathbf{B}_2,\dots,\mathbf{B}_n$ at respective temperatures $T_1 = (k_B\beta_1)^{-1}$, $T_2 = (k_B\beta_2)^{-1},\dots,T_n = (k_B\beta_n)^{-1}$. Let x be the phase-space point of system $\mathbf{S}$ and λ be a set of external parameters such as the volume of a gas or the frequency of optical tweezers. We control the system from time 0 to τ with control protocol $\lambda(t)$. Let $x(t)$ be a trajectory of the system.

To formulate the stochastic dynamics, we discretize the time interval $[0,\tau]$ by dividing it into N small intervals with width $\Delta t := \tau/N$. The original continuous-time dynamics is recovered by taking the limit of $N\to\infty$ or equivalently $\Delta t \to 0$. Let $t = n\Delta t$ and $x_n := x(n\Delta t)$. We refer to "time t" as "time $t_n := n\Delta t$." Then trajectory $\{x(t')\}_{t'\in[0,t]}$ corresponds to $X_n := (x_0,x_1,\dots,x_n)$.

A control protocol $\lambda(t)$ can also be discretized. Let λ_n be the value of λ between $t_n = n\Delta t$ and $t_{n+1} = (n+1)\Delta t$, where it is assumed to be constant during this time interval (see Fig. 1). We denote the trajectory of λ from time 0 to t_n as $\Lambda_n := (\lambda_0,\lambda_1,\dots,\lambda_{n-1})$. Let $\lambda_{\rm int}$ be the value of parameter λ before time 0, which is not necessarily equal to λ_0 because we can switch the value of the parameter at time 0. We also denote the value of λ after time $t_N := \tau$ as $\lambda_{\rm fin}$, which is not necessarily equal to λ_N either (see also Fig. 1).

Let $P_n[x_n]$ be the probability distribution of x at time t_n. In particular, $P_0[x_0]$ is the initial distribution of x. The initial distribution can be chosen as a stationary distribution under external parameters $\lambda_{\rm int}$, as $P_s[x_0|\lambda_{\rm int}]$, which means $P_0[x_0] = P_s[x_0|\lambda_{\rm int}]$. We note that $P_s[x_0|\lambda_{\rm int}]$ is not necessarily a canonical distribution; it can be a nonequilibrium stationary distribution. Due to the causality, x_{n+1} is determined by X_n through the transition probability $P[x_{n+1}|X_n,\lambda_n]$, which depends on the external parameters at time t_n (i.e., λ_n). We note that $P[x_{n+1}|X_n,\lambda_n]$ represents the the probability of realizing x_{n+1} at time t_{n+1} under the condition that the trajectory of x up to time t_n is given by X_n. If the dynamics is Markovian, $P[x_{n+1}|X_n,\lambda_n]$ can be replaced by $P[x_{n+1}|x_n,\lambda_n]$.

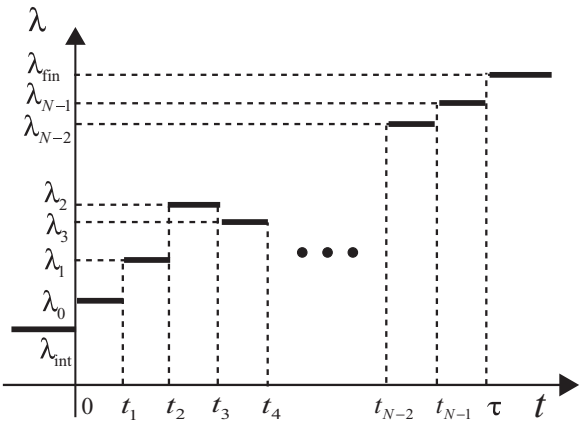

FIG. 1. Discretization of control protocol $\lambda(t)$.

The probability of trajectory X_n is then given by

$$P[X_n|\Lambda_n] = \prod_{k=0}^{n-1} P[x_{k+1}|X_k,\lambda_k]P_0[x_0] =: P[X_n], \qquad (1)$$

where we write $P[X_n|\Lambda_n]$ just as $P[X_n]$ for simplicity. We note that

$$P[X_n|x_0,\Lambda_n] = \prod_{k=0}^{n-1} P[x_{k+1}|X_k,\lambda_k] =: P[X_n|x_0] \qquad (2)$$

is the probability of trajectory X_n under the condition that the initial state is x_0 and the control protocol is Λ_n.

Let A be an arbitrary physical quantity that can depend on the trajectory X_N and protocol Λ_N. The ensemble average of this quantity is given by

$$\langle A\rangle = \int dX_N P[X_N|\Lambda_N]A[X_N,\Lambda_N], \qquad (3)$$

where $dX_N := \prod_{n=0}^{N} dx_n$.

B. Backward control

Before proceeding to the nonequilibrium equalities, we consider the stochastic dynamics with a backward control protocol. The backward control protocol means the time reversal of protocol Λ_N, which is formulated as follows. Let λ^* be the time reversal of λ; for example, if λ is a magnetic field, then $\lambda^* = -\lambda$. The time-reversed protocol of $\lambda(t)$ is then given by $\lambda^\dagger(t) := \lambda^*(\tau - t)$. The backward protocol can be discretized as $\Lambda_n^\dagger := (\lambda_{N-1}^*,\lambda_{N-2}^*,\dots,\lambda_{N-n-1}^*)$. We define $\lambda_n^\dagger := \lambda_{N-n-1}^*$, $\lambda_{\rm int}^\dagger := \lambda_{\rm fin}^*$, and $\lambda_{\rm fin}^\dagger := \lambda_{\rm int}^*$.

We consider the probability of realizing trajectory $x'(t)$ of the system with a backward control protocol. We define $x'_n := x'(n\Delta t)$ and $X'_n := (x'_0,x'_1,\dots,x'_N)$. We denote as $P_0^\dagger[x'_0]$ the

initial distribution of the backward processes. We stress that $P_0^\dagger[x_0']$ is not necessarily equal to the final distribution of the forward experiments. In fact, we can prepare a new state for the system to perform the backward experiments after the forward experiments. The probability distribution of trajectory X_n' with backward protocol is given by

$$P[X_N'|\Lambda_N^\dagger] = \prod_{k=0}^{N-1} P[x_k'|X_k',\lambda_k^\dagger]P_0^\dagger[x_0'] =: P^\dagger[X_N'], \quad (4)$$

where we write $P[X_N'|\Lambda_N^\dagger]$ as $P^\dagger[X_N']$ for simplicity. Correspondingly,

$$P[X_N'|x_0',\Lambda_N^\dagger] = \prod_{k=0}^{N-1} P[x_k'|X_k',\lambda_k^\dagger] =: P^\dagger[X_N'|x_0']. \quad (5)$$

In special cases, the backward trajectory X_N' is equal to the time reversal of the forward trajectory X_N. Let x^* be the time reversal of phase-space point x. For example, if $x = (\boldsymbol{r},\boldsymbol{p})$ with $\boldsymbol{r}$ and $\boldsymbol{p}$ being the position and the momentum, respectively, we have $x^* := (\boldsymbol{r}, -\boldsymbol{p})$. The time reversal of trajectory X_n is then given by $X_n^\dagger := (x_N^*, x_{N-1}^*, \ldots, x_{N-n}^*)$. With notation $x_n^\dagger := x_{N-n}^*$, we write $X_n^\dagger = (x_0^\dagger, x_1^\dagger, \ldots, x_n^\dagger)$. By substituting $x_n' = x_n^\dagger$ to Eqs. (4) and (5), we obtain the probability of realizing a backward trajectory under the backward protocol as

$$P^\dagger[X_N^\dagger] = \prod_{k=0}^{N-1} P[x_k^\dagger|X_k^\dagger,\lambda_k^\dagger]P_0^\dagger[x_0], \quad (6)$$

where the conditional probability under initial $x_0^\dagger$ is given by

$$P^\dagger[X_N^\dagger|x_0^\dagger] = \prod_{k=0}^{N-1} P[x_k^\dagger|X_k^\dagger,\lambda_k^\dagger]. \quad (7)$$

We note that $dX_N^\dagger = dX_N$ holds, because $dx_n = dx_n^*$.

C. Nonequilibrium equalities

We now discuss nonequilibrium equalities. Let $Q_i[X_N,\lambda_N]$ be the heat that is absorbed by the system from the ith heat bath satisfying $Q_i[X_N,\Lambda_N] = -Q_i[X_N^\dagger,\Lambda_N^\dagger]$. We write $Q_i[X_N,\lambda_N]$ simply as $Q_i[X_N]$ for simplicity. It has been established that the following equality is satisfied for stochastic thermodynamic systems [11,12,16,25]:

$$\frac{P^\dagger[X_N^\dagger|x_0^*]}{P[X_N|x_0]} = \exp\left(\sum_i \beta_i Q_i[X_N]\right), \quad (8)$$

which is referred to as the detailed fluctuation theorem (or the transient fluctuation theorem). This is the starting point of our research. We can rewrite Eq. (8) as

$$\frac{P^\dagger[X_N^\dagger]}{P[X_N]} = e^{-\sigma[X_N]}, \quad (9)$$

where

$$\sigma[X_N] := -\ln P_0^\dagger[x_0^\dagger] + \ln P_0[x_0] - \sum_i \beta_i Q_i[X_N], \quad (10)$$

which is called the entropy production along trajectory X_N.

Various proofs of the detailed fluctuation theorem [Eqs. (8) and (9)] for stochastic systems have been presented, for example, in Refs. [11,12,25] for the Markovian stochastic dynamics and in Ref. [29] for non-Markovian Langevin systems. A proof of Eqs. (8) and (9) has also been given in Ref. [16] for the situations in which the total system including heat baths is treated as a Hamiltonian system and the initial states of the heat baths in the forward and backward processes are the canonical distributions. This proof can confirm the physical validity of the detailed fluctuation theorem even for the non-Markovian dynamics with multiple heat baths, as the stochastic dynamics can be reproduced as that of a partial system of the total Hamiltonian system including the heat baths. We also note that several equalities that are similar but not equivalent to Eqs. (8) and (9) have been derived for different situations. For example, the transient fluctuation theorem has been discussed for dynamical systems in Ref. [20]. The fluctuation theorem for nonequilibrium steady states has been discussed for stochastic systems [10,13] and dynamical systems [7,8].

From the detailed fluctuation theorem (9), we can show Crooks's fluctuation theorem as follows. We denote as $P[\sigma]$ the probability of finding the entropy production σ in the forward processes, satisfying

$$P[\sigma] = \int \delta(\sigma - \sigma[X_N])P[X_N]dX_N, \quad (11)$$

where $\delta(\cdot)$ is the delta function. On the other hand, let $P^\dagger[\sigma]$ be the probability of obtaining σ in the backward processes, satisfying

$$P^\dagger[\sigma] = \int \delta(\sigma - \sigma[X_N'])P^\dagger[X_N']dX_N'. \quad (12)$$

By using the detailed fluctuation theorem (9) and equality $\sigma[X_N] = -\sigma[X_N^\dagger]$, we obtain Crooks's fluctuation theorem

$$\frac{P^\dagger[-\sigma]}{P[\sigma]} = e^{-\sigma}. \quad (13)$$

The detailed fluctuation theorem (9) or Crooks's fluctuation theorem (13) leads to the integral fluctuation theorem

$$\langle e^{-\sigma}\rangle = 1, \quad (14)$$

where the ensemble average $\langle\cdots\rangle$ is taken over all trajectories under forward protocol [see Eq. (3)]. From the concavity of the exponential function, we obtain

$$\langle\sigma\rangle \geqslant 0, \quad (15)$$

which is an expression of the second law of thermodynamics: The ensemble-averaged entropy production is non-negative. By taking the ensemble average of the logarithm of both sides of Eq. (9), we have

$$\langle\sigma\rangle = \int dX_N P[X_N] \ln \frac{P[X_N]}{P^\dagger[X_N^\dagger]}, \quad (16)$$

which we will refer to as the Kawai-Parrondo-Broeck (KPB) equality [30,31]. The right-hand side of Eq. (16) is the Kullback-Leibler divergence (or the relative entropy) of $P[X_N]$ and $P^\dagger[X_N^\dagger]$, which is always positive. Therefore, Eq. (16) reproduces inequality (15).

If the probability distribution of σ is Gaussian, the cumulant expansion of Eq. (3) leads to a variant of the fluctuation-dissipation relation

$$\langle\sigma\rangle = \tfrac{1}{2}(\langle\sigma^2\rangle - \langle\sigma\rangle^2), \tag{17}$$

which indicates that $\langle\sigma\rangle$ is determined by the fluctuation of σ. Equality (17) is an expression of the fluctuation-dissipation theorem of the first kind, which gives a special case of the Green-Kubo formula [20].

In the case of an isothermal process with a single heat bath, the entropy production reduces to

$$\sigma[X_N] = \beta(W[X_N] - \Delta F), \tag{18}$$

where $W[X_N]$ is the work performed on the system during the process, and ΔF is the difference of the free energies for the initial and final Hamiltonians (see also the Appendix for details). Under this situation, Eq. (14) leads to

$$\langle e^{-\beta W}\rangle = e^{-\beta\Delta F}, \tag{19}$$

which is the Jarzynski equality [9]. The second law of thermodynamics then reduces to

$$\langle W\rangle \geqslant \Delta F. \tag{20}$$

III. MEASUREMENT

In this section we formulate and investigate the effect of measurements on nonequilibrium dynamics.

A. Classical measurement and mutual information

In this subsection we review the general framework of a measurement on a probabilistic variable, which can be applied to a broad class of measurements on classical systems.

Let x be an arbitrary probability variable of a measured system whose distribution is $P[x]$. We perform a measurement on it and obtain outcome y, which is also a probability variable. The error of the measurement can be characterized by a conditional probability $P[y|x]$, which describes the probability of obtaining outcome y under the condition that the true value of the measured system is x. We note that $\sum_y P[y|x] = 1$ for all x, where we note that the sum should be replaced by the integral if y is a continuous variable. If the measurement is error free, $P[y|x]$ is given by the delta function or the Kronecker's delta. We assume that $P[y|x]$ is independent of the probability distribution $P[x]$; in other words, the error is independent of the state preparation of the measured system. The joint probability of x and y is given by $P[x,y] = P[y|x]P[x]$, and the probability of obtaining y by $P[y] = \sum_x P[x,y]$. The probability of realizing x under the condition that the measurement outcome is y, denoted as $P[x|y]$, is given by Bayes' theorem:

$$P[x|y] = \frac{P[y|x]P[x]}{P[y]}. \tag{21}$$

We next discuss the information contents related to the measurement [105–107]. The Shannon information contents of the probability variables are given by

$$H_x := -\sum_x P[x]\ln P[x],\ H_y := -\sum_y P[y]\ln P[y], \tag{22}$$

which characterize the randomnesses of x and y, respectively. On the other hand, the mutual information content $\langle I\rangle$ between x and y is given by

$$\langle I\rangle := \sum_{xy} P[x,y]I[x:y], \tag{23}$$

where

$$I[x:y] := \ln\frac{P[y|x]}{P[y]}. \tag{24}$$

In this paper we also call $I[x:y]$ the mutual information. We note that $I[x:y] = I[y:x]$ holds due to Bayes' theorem (21).

The mutual information $\langle I\rangle$ measures the amount of information obtained by the measurement. It is known that

$$0 \leqslant \langle I\rangle \leqslant H_x, \quad 0 \leqslant \langle I\rangle \leqslant H_y. \tag{25}$$

If the measurement is error free, $\langle I\rangle = H_x = H_y$ holds.

B. Measurements on nonequilibrium dynamics

We next formulate multiple measurements on nonequilibrium dynamics and discuss the properties of the mutual information obtained by the measurements.

Let y_n be the outcome at time $t_n := n\Delta t$. In this section we assume the following:

(1) The error of the measurement at time t_n is characterizes by $P[y_n|X_n]$, where y_n can depend on the trajectory of the system before t_n due to the causality. Here we assumed that the property of the measurement error at time t_n does not explicitly depend on Y_{n-1} or $P[X_n]$. This assumption is also justified in many real experimental situations.

(2) The unconditional probability distribution of X_n, $P[X_n]$, is not affected by the back-action of the measurement. Since the system is classical, this assumption is justified for many real systems such as colloidal particles and macromolecules.

If $P[y_n|X_n] = P[y_n|x_n]$, we call the measurement Markovian, which means that the outcome is determined only by the system's state immediately before the measurement. This condition is satisfied if the measurements can be performed in a time interval that is sufficiently shorter than the shortest time scale Δt of the system. We note that the Markovness of the measurement is independent of that of the dynamics.

We assume that the measurements are performed at times $t_{n_1}, t_{n_2}, \ldots, t_{n_M}$, where $0 \leqslant n_1 < n_2 < \cdots < n_M \leqslant N$. If $n_1 = 0$, $n_2 = 1$, $n_3 = 2, \ldots, n_{N+1} = N$ hold, the measurement is time continuous in the limit of $\Delta t \to 0$, because the measurements are performed at all times.

We write as Y_n the set of measurement outcomes that are obtained up to time t_n, i.e., $Y_n := (y_{n_1}, y_{n_2}, \ldots, y_{[n]})$, where $[n]$ is the maximum n_k satisfying $n_k \leqslant n$. If the measurement is continuous, then $Y_n = (y_0, y_1, \ldots, y_n)$.

We define

$$P_c[Y_n|X_n] := \prod_{k=1}^{M'} P[y_{n_k}|X_{n_k}], \tag{26}$$

where M' is the maximum integer satisfying $n_{M'} \leqslant n$. Without feedback, Eq. (26) defines the conditional probability of obtaining outcomes Y_n under the condition of X_n, while, with feedback, this interpretation of Eq. (26) is not necessarily correct, as shown in the next section. To explicitly demonstrate this point and to distinguish $P_c[Y_n|X_n]$ from the usual conditional probability, we put the subscript c. Then the joint distribution of X_n and Y_n is given by

$$P[X_n,Y_n] = P_c[Y_n|X_n]P[X_n]. \tag{27}$$

The probability of obtaining outcomes Y_n is given by

$$P[Y_n] = \int dX_n P[X_n,Y_n] = \prod_{k=1}^{M'} P[y_{n_k}|Y_{n_k-1}], \tag{28}$$

where the two equalities are just identities known in probability theory. We also note that

$$\begin{aligned} P[y_n|X_n,Y_{n-1}] &:= \frac{P[Y_n|X_n]}{P[Y_{n-1}|X_n]} \\ &= P[y_n|X_n], \end{aligned} \tag{29}$$

which is, in fact, independent of Y_{n-1}.

We then discuss the mutual information obtained by multiple measurements on nonequilibrium dynamics. Suppose that we obtain measurement outcomes Y_{n-1} until time t_{n-1}. If we perform another measurement at time t_n and obtain outcome y_n, we obtain the mutual information between y_n and X_n under the condition that we have obtained Y_{n-1}:

$$\begin{aligned} I[y_n : X_n|Y_{n-1}] &:= \ln \frac{P[y_n|X_n,Y_{n-1}]}{P[y_n|Y_{n-1}]} \\ &= \ln \frac{P[y_n|X_n]}{P[y_n|Y_{n-1}]}, \end{aligned} \tag{30}$$

where we used Eq. (29). We note that, if the measurement is Markovian, $I[y_n : X_n|Y_{n-1}]$ reduces to $I[y_n : x_n|Y_{n-1}]$. The mutual information $\langle I[y_n : X_n|Y_{n-1}]\rangle := \int dX_n dY_n P[X_n,Y_n] I[y_n : X_n|Y_{n-1}]$ is called the transfer entropy, which describes the information flow from the system to the outcome as discussed in Ref. [107]. We note that the same quantity has been discussed in Ref. [62].

We denote as I_c the sum of these mutual information contents obtained by multiple measurements, that is,

$$\begin{aligned} I_c[X_n : Y_n] &:= \sum_{k=1}^{M'} I[y_{n_k} : X_{n_k}] \\ &= \ln \frac{P[Y_n|X_n]}{P[Y_n]}, \end{aligned} \tag{31}$$

where we used Eq. (28). From Eq. (31), we find that $I_c[Y_n : X_n]$ equals the mutual information between trajectories X_n and Y_n defined as $I[Y_n : X_n] := \ln(P[Y_n|X_n]/P[Y_n])$. In the presence of feedback control, however, this is not true (i.e., $I_c \neq I$), as we will see later.

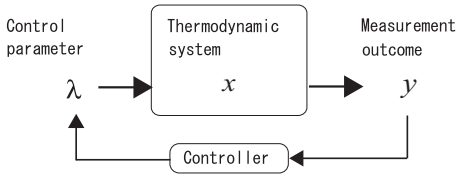

FIG. 2. Feedback control on nonequilibrium dynamics. The control parameter is denoted as λ, the point of the phase space of the system as x, and the outcome of measurement on the system as y. Parameter λ depends on y through the real-time feedback control.

IV. FEEDBACK CONTROL

In this section we formulate feedback control on nonequilibrium dynamics.

A. Formulation

Feedback control implies that protocol Λ_N depends on measurement outcomes Y_N (see Fig. 2). On the other hand, without feedback control, control protocols are predetermined and independent of the measurement outcomes, as is the case for the setup of the original fluctuation theorem and Jarzynski equality.

When the system is subject to feedback control, λ_n can depend on measurement outcomes that are obtained until t_n, while λ_n cannot depend on any measurement outcome that is obtained after time t_n due to causality. We introduce the notation $\lambda_n(Y_n)$, which means that the value of λ at time t_n is determined by Y_n. We write $\Lambda_n(Y_{n-1}) := [\lambda_0(Y_0),\lambda_1(Y_1),\ldots,\lambda_{n-1}(Y_{n-1})]$.

If λ_n depends only on y_n as $\lambda_n(y_n)$, the feedback protocol is called Markovian. We note that the Markovian quality of feedback is independent of that of the dynamics or measurements. The Markovian feedback control is realized when the delay time of feedback is sufficiently smaller than the smallest time scale Δt of the dynamics.

B. Overdamped Langevin system

As a simple illustrative example, we discuss an overdamped Langevin system, whose equation of motion is given by

$$\eta \frac{dx(t)}{dt} = -\frac{\partial V(x,\lambda)}{\partial x} + f(\lambda) + \sqrt{2\eta k_B T}\xi(t), \tag{32}$$

where η is the friction constant, $V(x,\lambda)$ is an external potential, $f(\lambda)$ is an external nonconservative force, and $\xi(t)$ is the Gaussian white noise satisfying $\langle \xi(t)\xi(t')\rangle = \delta(t-t')$. The detailed fluctuation theorem (8) is still satisfied in the presence of a nonconservative force that violates the detailed balance, because Eq. (8) can be derived from the local transition rate of the stochastic dynamics and is independent whether there is a global potential or not. We assume that the measurement is time continuous and Markovian:

$$y_n = x_n + \frac{\Delta R_n}{\Delta t}, \tag{33}$$

where ΔR_n is a white Gaussian noise with $\langle \Delta R_n \Delta R_{n'} \rangle = R\delta_{nn'}\Delta t$ $(R > 0)$. The conditional probability of obtaining outcome y_n is given by

$$P[y_n|x_n] \propto \exp\left[-\frac{\Delta t}{2R}(y_n - x_n)^2\right]. \tag{34}$$

The feedback protocol can be written as $\lambda_n(y_0, y_1, \ldots, y_n)$ in general. The work performed on the system is then given by

$$\begin{aligned} W_n &:= V(x_n, \lambda_{n+1}) - V(x_n, \lambda_n) \\ &= \frac{\partial V}{\partial \lambda}\Delta\lambda_n + O(\Delta t), \end{aligned} \tag{35}$$

where

$$\Delta\lambda_n := \lambda_{n+1}(y_0, y_1, \ldots, y_{n+1}) - \lambda_n(y_0, y_1, \ldots, y_n). \tag{36}$$

In particular, if the feedback is Markovian, λ_n is given by $\lambda_n(y_n)$. Then $\Delta\lambda_n = \lambda_{n+1}(y_{n+1}) - \lambda_n(y_n)$ can be written as

$$\Delta\lambda_n = \frac{\partial\lambda}{\partial t}\Delta t + \frac{\partial\lambda}{\partial y}\Delta y_n + \frac{1}{2}\frac{\partial^2\lambda}{\partial y^2}\Delta y_n^2, \tag{37}$$

where $\Delta y_n := y_{n+1} - y_n$. The first term on the right-hand side of Eq. (37) arises from a change in λ by the pre-fixed protocol, while the second and third terms are induced by the feedback control.

We next consider the Kalman filter and the optimal control. As a special case of Eq. (32), we consider a discretized linear Langevin equation in the Itô form:

$$\eta(x_{n+1} - x_n) = -Kx_n\Delta t + \lambda_n\Delta t + \sqrt{2\eta k_B T}\,\Delta W_n, \tag{38}$$

where K is a positive constant and λ_n is a control parameter. If the initial distribution of x_0 is Gaussian, the distribution of x_n remains Gaussian with Eq. (38). In this case the obtained mutual information by measurement (33) at time t_n is given by

$$\langle I_c[x_n : y_n] \rangle = \frac{1}{2}\ln\left(1 + \frac{S_n}{R}\Delta t\right) = \frac{S_n}{2R}\Delta t + o(\Delta t), \tag{39}$$

where $S_n := \langle x_n^2 \rangle - \langle x_n \rangle^2$. Therefore, the total mutual information

$$\langle I_c \rangle = \lim_{N\to\infty}\sum_{n=0}^{N}\frac{S_n}{2R}\Delta t \tag{40}$$

can converge, while the measurement is continuous.

We consider the Kalman filter on Eq. (38) with measurement (33). The Kalman filter is a standard method to construct the optimal estimator of x_n, denoted as $\hat{x}_n$, in terms of the mean-square error. From measurement outcomes Y_n, $\hat{x}_n$ is obtained as the solution to the following simultaneous differential equations [108]:

$$\hat{x}_{n+1} - \hat{x}_n = -\frac{K\hat{x}_n + \lambda_n}{\eta}\Delta t + \frac{A_n}{R}(y_n - \hat{x}_n)\Delta t, \tag{41}$$

$$A_{n+1} - A_n = \left(\frac{-2KA_n + 2k_B T}{\eta} - \frac{A_n}{R}\right)\Delta t, \tag{42}$$

where A_n is a time-dependent real number and Eq. (42) is a discretized version of the Riccati equation. By using the Kalman estimator $\hat{x}_n$, the optimal control protocol [109] is given by

$$\lambda_n = -C_n\hat{x}_n, \tag{43}$$

where C_n is a predetermined constant depending on the target of the optimal control. We note that the optimal control is a non-Markovian control as $\lambda_n = \lambda_n(Y_n)$, because we use all of $Y_n = (y_0, y_1, \ldots, y_n)$ to calculate $\hat{x}_n$. The generalized Jarzynski equality for this situation has been discussed in Ref. [68].

C. Probability distributions with feedback

We discuss the probability distributions with feedback control in general. Under the condition that we fix control protocol $\Lambda_N(Y_N)$ with Y_N being fixed, the conditional probability of realizing X_N is given by

$$P[X_n|\Lambda_n(Y_{n-1})] = P_0[X_0]\prod_{k=0}^{k-1}P[x_{k+1}|X_k, \lambda_k(Y_k)], \tag{44}$$

which corresponds to Eq. (1). We note that, in the expression in Eq. (44), we do not omit the notation $\Lambda_N(Y_{N-1})$ because its Y_{N-1} dependence is crucial. We also write

$$P[X_n|x_0, \Lambda_n(Y_{n-1})] := \prod_{k=0}^{k-1}P[x_{k+1}|X_k, \lambda_k(Y_k)]. \tag{45}$$

On the other hand, along the trajectory X_n, the conditional probability of obtaining outcome y_n at time t_n is written as $P[y_n|X_n]$. We then define

$$P_c[Y_n|X_n] := \prod_{k=0}^{n-1}P[y_k|X_k], \tag{46}$$

which is to be compared with Eq. (26).

We then obtain the joint probability distribution of X_n and Y_n with feedback control as

$$\begin{aligned} P[X_n, Y_n] &= \prod_{k=0}^{n-1}P[y_{k+1}|X_{k+1}]P[x_{k+1}|X_k, \lambda_k(Y_{k-1})] \\ &= P_c[Y_n|X_n]P[X_n|\Lambda_n(Y_{n-1})]. \end{aligned} \tag{47}$$

We can check that

$$\int P[X_n, Y_n]dX_n dY_n = 1, \tag{48}$$

by integrating X_n and Y_n in Eq. (47) in the order of $y_n \to x_n \to y_{n-1} \to x_{n-1} \to \cdots \to y_1 \to x_1 \to y_0 \to x_0$, where the causality of measurements and feedback play crucial roles.

The marginal distributions are given by

$$P[X_n] = \int P[X_n, Y_n]dY_n, \; P[Y_n] = \int P[X_n, Y_n]dX_n, \tag{49}$$

and the conditional distributions by

$$P[X_n|Y_n] = \frac{P[X_n, Y_n]}{P[Y_n]}, \quad P[Y_n|X_n] = \frac{P[X_n, Y_n]}{P[X_n]}. \tag{50}$$

We stress that, in the presence of feedback control,

$$P[Y_n|X_n] \neq P_c[Y_n|X_n] \tag{51}$$

in general, because protocol Λ_N depends on Y_{N-1}. On the other hand, without feedback control, $P[Y_n|X_n] = P_c[Y_n|X_n]$ holds because $P[X_n]$ is simply given by $P[X_n|\Lambda_n]$ with Λ_n being independent of Y_n.

The ensemble average of a probability variable $A[X_n,Y_n]$ is given by

$$\langle A\rangle := \int A[X_n,Y_n]P[X_n,Y_n]dX_ndY_n, \tag{52}$$

and the conditional average under the condition of Y_n is given by

$$\langle A\rangle_{Y_n} := \int A[X_n,Y_n]P[X_n|Y_n]dX_n. \tag{53}$$

Equation (29) still holds in the presence of feedback control:

$$\begin{aligned} P[y_n|X_n,Y_{n-1}] &:= \frac{P[X_n,Y_n]}{P[X_n,Y_{n-1}]} \\ &= \frac{P_{\rm c}[Y_n|X_n]P[X_n|\Lambda_n(Y_{n-1})]}{P_{\rm c}[Y_{n-1}|X_{n-1}]P[X_n|\Lambda_n(Y_{n-1})]} \\ &= P[y_n|X_n]. \end{aligned} \tag{54}$$

We note that Eq. (28) also holds with feedback control.

We then define the mutual information (or the transfer entropy) in the same way as in the case without feedback control:

$$\begin{aligned} I_{\rm c}[Y_n:X_n] &:= \sum_{k=1}^{M'} I[y_{n_k}:X_{n_k}|Y_{n_k-1}] \\ &= \ln\frac{P_{\rm c}[Y_n|X_n]}{P[Y_n]}. \end{aligned} \tag{55}$$

In the presence of feedback control, $I_{\rm c}[Y_n:X_n]$ does not equal the mutual information between trajectories X_n and Y_n defined as $I[Y_n:X_n] := \ln(P[Y_n|X_n]/P[Y_n])$, because $P_{\rm c}[Y_n|X_n] \neq P[Y_n|X_n]$. Intuitively speaking, $I_{\rm c}$ characterizes only the correlation between X_n and Y_n due to the measurements, while I involves the correlation due to the feedback control. Note that $I_{\rm c}$ is a more important quantity than I, because $I_{\rm c}$ has a clear information-theoretic significance: $I_{\rm c}$ is the information that we obtain by measurements. We also note that, in the case of a single measurement and feedback, $I_{\rm c} = I$ always holds.

We also note that an identity similar to the integral fluctuation theorem holds for $I_{\rm c}$:

$$\langle e^{-I_{\rm c}}\rangle = 1, \tag{56}$$

because

$$\begin{aligned} \langle e^{-I_{\rm c}}\rangle &= \int dX_NdY_N\frac{P[Y_N]}{P_{\rm c}[Y_N|X_N]}P[X_N,Y_N] \\ &= \int dX_NdY_NP[Y_N]P[X_N|\Lambda_N(Y_{N-1})] = 1. \end{aligned} \tag{57}$$

D. Detailed fluctuation theorem for a fixed control protocol

If we fix control protocol $\Lambda_N(Y_N)$ with Y_N being fixed, then the detailed fluctuation theorem (8) still holds:

$$\frac{P[X_N^\dagger|x_0^*,\Lambda_N(Y_{N-1})^\dagger]}{P[X_N|x_0,\Lambda_N(Y_{N-1})]} = \exp\left\{\sum_i \beta_i Q_i[X_N,\Lambda_N(Y_{N-1})]\right\}, \tag{58}$$

where

$$\Lambda_N(Y_N)^\dagger := [\lambda_{N-1}(Y_{N-1})^*,\dots,\lambda_0(Y_0)^*]. \tag{59}$$

The left-hand side of Eq. (58) corresponds to the following forward and backward experiments. We first perform forward experiments many times with feedback control and choose the subensemble in which the measurement outcomes are given by Y_{N-1}. Within this subensemble, the ratio of trajectory X_N is given by $P[X_N|x_0,\Lambda_N(Y_{N-1})]$ under the condition of initial x_0. We next perform backward experiments with protocol $\Lambda_N(Y_{N-1})^\dagger$, where Y_{N-1} was chosen in the forward experiments. We stress that we do not perform any feedback in the backward experiments: $\Lambda_N(Y_{N-1})^\dagger$ is just the time reversal of $\Lambda_N(Y_{N-1})$. We then obtain $P[X_N^\dagger|x_0^*,\Lambda_N(Y_{N-1})^\dagger]$ as the ratio of trajectory $X_N^\dagger$, under the condition of initial $x_0^\dagger$ in the backward experiments. The original detailed fluctuation theorem (8) can straightforwardly be applied to this subensemble corresponding to Y_{N-1} because we have a unique control protocol in the subensemble, and therefore we obtain Eq. (58).

Let the initial distribution of the backward experiments be $P_0^\dagger[x_0^\dagger|Y_N]$, which in general depends on the measurement outcomes in the forward experiments. A natural choice of $P_0^\dagger[x_0^\dagger|Y_N]$ is a stationary state $P_{\rm s}[x_0^\dagger|\lambda(Y_N)^*]$. Then we have

$$\frac{P[X_N^\dagger|\Lambda_N(Y_N)^\dagger]}{P[X_N|\Lambda_N(Y_N)]} = \exp\{-\sigma[X_N,\Lambda_N(Y_N)]\}, \tag{60}$$

where

$$\begin{aligned} \sigma[X_N,\Lambda_N(Y_N)] &:= -\ln P_0^\dagger[x_0^\dagger|Y_N] + \ln P_0[x_0] \\ &\quad - \sum_i \beta_i Q_i[X_N,\Lambda_N(Y_{N-1})]. \end{aligned} \tag{61}$$

If there is a single heat bath and the initial distributions of the forward and backward experiments are given by the canonical distributions, then the entropy production reduces to

$$\sigma[X_N,\Lambda_N(Y_N)] = \beta\{W[X_N,\Lambda_N(Y_N)] - \Delta F[Y_N]\}, \tag{62}$$

where the free-energy difference can depend on the measurement outcomes as $\Delta F[Y_N] := F[\lambda_{\rm fin}(Y_N)] - F(\lambda_{\rm int})$.

V. NONEQUILIBRIUM EQUALITIES WITH FEEDBACK CONTROL

We now discuss the main results of this paper. We derive the two types of the generalized nonequilibrium equalities with feedback control in Secs. V A and V B, respectively. The former generalization involves the mutual information, while the latter involves the efficacy of feedback control.

A. Generalized fluctuation theorem with mutual information

To derive a generalized detailed fluctuation theorem, we first formulate the relevant backward probabilities. We consider the following type of "backward probability distribution":

$$P^\dagger[X_N^\dagger,Y_N] := P[X_N^\dagger|\Lambda_N(Y_{N-1})^\dagger]P[Y_N], \tag{63}$$

which satisfies

$$\int P^\dagger[X_N^\dagger,Y_N]dX_N^\dagger dY_N^\dagger = 1. \tag{64}$$

Definition (63) has a clear operational meaning. Suppose that we perform a forward experiment with feedback and obtain

outcome Y_N. We then perform a backward experiment with protocol $\Lambda_N(Y_{N-1})^\dagger$. We repeat this set of the forward and backward experiments many times, and calculate the fractions of (X_N,Y_N) and $(X_N^\dagger,Y_N)$, which, respectively, give $P[X_N,Y_N]$ and $P^\dagger[X_N^\dagger,Y_N]$.

Noting Eq. (47) and the definition of the mutual information (55), we obtain a generalized detailed fluctuation theorem with feedback control:

$$\frac{P^\dagger[X_N^\dagger,Y_N]}{P[X_N,Y_N]} = \exp\left(-\sigma[X_N,\Lambda_N(Y_N)] - I_c[X_N:Y_N]\right), \tag{65}$$

where the effect of feedback control is involved by the term of the mutual information that is obtained in the forward experiments. We stress that, to obtain Eq. (65), we do not perform feedback control in the backward experiments. We just reverse forward protocol as Eq. (59) in the backward experiments. The same result for a special case was obtained in Ref. [71]. The investigation of the detailed fluctuation theorem in the situations in which feedback control is also performed in the backward processes [70] is an interesting future challenge. Such situations would be relevant to, for example, autonomous systems consisting of the controlled system and the controller, in which feedback control should also be needed for the backward processes. We can expect that the backward processes with feedback control can be used to characterize the reversibility of the autonomous systems.

From the generalized detailed fluctuation theorem (65), we obtain a generalized integral fluctuation theorem [67]:

$$\langle e^{-\sigma-I_c}\rangle = 1. \tag{66}$$

Due to the concavity of the exponential function, we obtain a generalized second law of thermodynamics [67,100]:

$$\langle\sigma\rangle \geqslant -\langle I_c\rangle, \tag{67}$$

which means that the entropy production can be negative due to the effect of feedback control (or due to the action of Maxwell's demon), and that the lower bound of the entropy production is bounded by the mutual information $\langle I_c\rangle$.

The reason why the entropy production can be negative is that one can rectify the thermal fluctuations by feedback control. This negative entropy production is compensated for by the excess entropy production in the demon or the feedback controller [102], and therefore the entropy production in the total system consisting of the demon and the information heat engine is consistent with the second law of thermodynamics. The key feature of feedback control is that it enables us to control the entropy production of a partial system by utilizing the mutual information beyond the limitation of the conventional thermodynamics. Inequality (67) identifies the lower bound of the entropy production with feedback control, which plays a role parallel to the conventional second law of thermodynamics that gives the lower bound of zero in the absence of feedback control. Therefore, inequality (67) is regarded as a generalization of the second law of thermodynamics that can be applied to feedback-controlled processes.

We also obtain, by taking the ensemble average of the logarithm of the both sides of Eq. (65), that

$$\langle\sigma\rangle + \langle I_c\rangle = \int dX_N dY_N P[X_N,Y_N] \ln \frac{P[X_N,Y_N]}{P^\dagger[X_N^\dagger,Y_N]}, \tag{68}$$

which is a generalization of the KPB equality (16). We note that the right-hand side of Eq. (68) is positive because it is the Kullback-Leibler divergence between two probability distributions $P[X_N,Y_N]$ and $P^\dagger[X_N^\dagger,Y_N]$; thus, inequality (67) is reproduced. We note that equality in Eq. (67) is achieved if and only if $\sigma + I_c$ does not fluctuate, or equivalently, if

$$P[X_N,Y_N] = P^\dagger[X_N^\dagger,Y_N] \tag{69}$$

holds, which implies the reversibility with feedback control [75]. The more the probability distribution of the forward processes with feedback is different from that of the backward processes without feedback, the more $\langle\sigma\rangle$ is different from $-\langle I_c\rangle$.

If the joint distribution of σ and I_c is Gaussian, we have a generalized fluctuation-dissipation theorem from the second cumulant of Eq. (66):

$$\langle\sigma + I_c\rangle = \tfrac{1}{2}[\langle(\sigma+I_c)^2\rangle - \langle\sigma+I_c\rangle^2], \tag{70}$$

which suggests that there is a tradeoff relation between the entropy production and the mutual information.

For the case in which $\sigma = \beta(W-\Delta F)$, Eq. (65) leads to a generalized Jarzynski equality:

$$\langle e^{-\beta(W-\Delta F)-I_c}\rangle = 1, \tag{71}$$

and inequality (67) leads to

$$\langle W\rangle \geqslant \Delta F - k_B T\langle I_c\rangle. \tag{72}$$

We note that Eq. (71) and inequality (72) are the generalizations of the results obtained in Refs. [67,71]. By defining $W_{\rm ext} := -W$ and setting $\Delta F = 0$, we can rewrite inequality (72) as

$$\langle W_{\rm ext}\rangle \leqslant k_B T\langle I_c\rangle, \tag{73}$$

which implies that we can extract a positive work up to the term that is equal to the mutual information multiplied by $k_B T$, from a thermodynamic cycle with a single heat bath with the assistance of feedback control or Maxwell's demon. The mutual information can be used as a "resource" of the work or the free energy. In the case of the Szilard engine, $\langle I_c\rangle = \langle I\rangle = \ln 2$ and $\langle W_{\rm ext}\rangle = k_B T\ln 2$ hold, and therefore the equality in Eq. (73) is achieved. In fact, in the Szilard engine, $\sigma + I = \beta(W-\Delta F) + I$ does not fluctuate, but is zero for both outcomes "left" and "right."

We note that, to obtain Eq. (66) or (71) experimentally or numerically, the condition of $P_c[Y_N|X_N] \neq 0$ needs to be satisfied for all (X_N,Y_N). To explicitly see this, we write $P_c[Y_N|X_N] =: \varepsilon > 0$. We then obtain

$$P[X_N,Y_N]e^{-\sigma-I_c} = \varepsilon P[X_N]\frac{1}{\varepsilon}e^{-\sigma+\ln P[Y_N]}, \tag{74}$$

which does not converge to zero with the limit of $\varepsilon \to 0$. On the other hand, in real experiments or numerical simulations, the events with $P[X_N,Y_N] = 0$ never occur. Therefore, if $P_c[Y_N|X_N] = 0$ holds for some (X_N,Y_N), the terms associated

with zero-probability events make nonzero contributions to Eq. (66) and (71); in such cases, we cannot obtain Eq. (66) or (71) experimentally or numerically. On the contrary,

$$P[X_N,Y_N]I_{\rm c}^n = \varepsilon P[X_N]\left(\ln\frac{\varepsilon}{P[Y_N]}\right)^n \tag{75}$$

converges to zero for all $n = 1,2,\dots,$ in the limit of $\varepsilon \to 0$. Therefore, we can find $\langle I_{\rm c}^n\rangle$ experimentally and numerically even if $P_{\rm c}[Y_N|X_N] = 0$ for some (X_N,Y_N), and also obtain Eqs. (68) and (70) and inequalities (67), (72), and (73).

B. Generalized fluctuation theorem with efficacy parameter

We next derive a different type of nonequilibrium equality. In this subsection we assume that the measurements are Markovian (i.e., $P[y_n|X_n] = P[y_n|x_n]$ holds). We perform forward experiments with measurements at times $t_{n_1},t_{n_2},\dots,t_{n_M}$ with feedback control, and perform backward experiments without feedback but only with measurements at times $t_{N-n_M},t_{N-n_{M-1}},\dots,t_{N-n_1}$.

Let $Y'_N := (y'_{N-n_M},y'_{N-n_{M-1}},\dots,y'_{N-n_1})$ be the measurement outcomes in the backward measurements. Then the probability of obtaining Y'_N under the condition of $X^\dagger_N$ is given by

$$P_{\rm c}[Y'_N|X^\dagger_N] := \prod_{k=1}^{M} P[y'_{N-n_k}|x^\dagger_{N-n_k}]. \tag{76}$$

Therefore, the probability of obtaining Y'_N under protocol $\Lambda(Y_N)^\dagger$ is

$$P[Y'_N|\Lambda_N(Y_{N-1})^\dagger] = \int P_{\rm c}[Y'_N|X^\dagger_N]P[X^\dagger_N|\Lambda_N(Y_{N-1})^\dagger]dX^\dagger_N, \tag{77}$$

which is normalized as

$$\int P[Y'_N|\Lambda_N(Y_{N-1})^\dagger]dY'_N = 1, \tag{78}$$

where the probability variable Y'_N is independent of Y_N.

We then consider the time-reversed sequence of Y_N. Let y_n^* be the time reversal of y_n; for example, if we measure the momentum, then $y_n^* = -y_n$. We write $Y^\dagger_N := (y^*_{N-n_M},y^*_{N-n_{M-1}},\dots,y^*_{N-n_1})$. The probability of $Y'_N = Y^\dagger_N$ in the backward experiments is given by

$$P[Y^\dagger_N|\Lambda_N(Y_{N-1})^\dagger] = \int P_{\rm c}[Y^\dagger_N|X^\dagger_N]P[X^\dagger_N|\Lambda_N(Y_{N-1})^\dagger]dX^\dagger_N, \tag{79}$$

which is the probability of obtaining the time-reversed outcomes by time-reversed measurements during the time-reversed protocol. We stress that

$$\int P[Y^\dagger_N|\Lambda_N(Y_{N-1})^\dagger]dY^\dagger_N \neq 1 \tag{80}$$

in general because $Y^\dagger_N$ is no longer independent of Y_{N-1}.

In the following, we assume that the measurements have the time-reversed symmetry

$$P[y_n^*|x_n^*] = P[y_n|x_n] \tag{81}$$

for all n, which leads to

$$P_{\rm c}[Y^\dagger_n|X^\dagger_n] = P_{\rm c}[Y_n|X_n]. \tag{82}$$

We then have the "renormalized" (or "coarse-grained") detailed fluctuation theorem [30,67]

$$\frac{P[Y^\dagger_N|\Lambda(Y_N)^\dagger]}{P[Y_N]} = e^{-\sigma'[Y_N]}, \tag{83}$$

where $\sigma'[Y_N]$ is the "renormalized" (or "coarse-grained") entropy production defined as

$$\begin{aligned}\sigma'[Y_N] &:= -\ln\langle e^{-\sigma}\rangle_{Y_N} \\ &= -\ln\int dX_N e^{-\sigma[X_N,\Lambda_N(Y_{N-1})]}P[X_N|Y_N].\end{aligned} \tag{84}$$

Equality (83) implies that the detailed fluctuation theorem retains its form under the coarse graining, if we introduce the appropriate coarse-grained entropy production. From the concavity of the exponential function, we obtain $\sigma'[Y_N] \leqslant \langle\sigma\rangle_{Y_N}$ and $\langle\sigma'\rangle \leqslant \langle\sigma\rangle$. The same result for a different setup has been obtained in Refs. [30,31].

The proof of Eq. (83) goes as follows. From the definition of $\sigma'[Y_N]$ and the detailed fluctuation theorem (58), we have

$$\begin{aligned}e^{-\sigma'[Y_N]} &= \int dX_N \frac{P[X^\dagger_N|\Lambda_N(Y_{N-1})^\dagger]}{P[X_N|\Lambda_N(Y_{N-1})]}P[X_N|Y_N] \\ &= \int dX_N \frac{P[X^\dagger_N|\Lambda_N(Y_{N-1})^\dagger]}{P[X_N|\Lambda_N(Y_{N-1})]}\frac{P[X_N,Y_N]}{P[Y_N]} \\ &= \frac{1}{P[Y_N]}\int dX_N P[X^\dagger_N|\Lambda_N(Y_{N-1})^\dagger]P_{\rm c}[Y_N|X_N] \\ &= \frac{1}{P[Y_N]}\int dX_N P[X^\dagger_N|\Lambda_N(Y_{N-1})^\dagger]P_{\rm c}[Y^\dagger_N|X^\dagger_N].\end{aligned} \tag{85}$$

In the last line, we used the time-reversal symmetry (82) of the measurements. By noting Eq. (79), we obtain Eq. (83).

We note that Eq. (83) holds regardless of the presence of feedback control. Without feedback control, Eq. (83) reduces to

$$\frac{P^\dagger[Y^\dagger_N]}{P[Y_N]} = e^{-\sigma'[Y_N]}. \tag{86}$$

By taking the ensemble average of both sides of Eq. (83) and noting that $\langle e^{-\sigma'}\rangle = \langle e^{-\sigma}\rangle$ holds, we obtain the second generalization of the integral fluctuation theorem [67]

$$\langle e^{-\sigma}\rangle = \gamma, \tag{87}$$

where γ is the efficacy parameter of feedback control defined as

$$\gamma := \int P[Y^\dagger_N|\Lambda_N(Y_{N-1})^\dagger]dY^\dagger_N, \tag{88}$$

which is the sum of probabilities of obtaining the time-reversed outcomes by the time-reversed measurements during the time-reversed protocols (see Fig. 3). If $\sigma = \beta(W-\Delta F)$ holds, Eq. (87) leads to the second generalization of the Jarzynski equality [67]:

$$\langle e^{-\beta(W-\Delta F)}\rangle = \gamma. \tag{89}$$

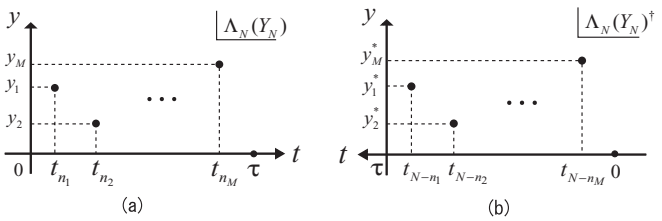

FIG. 3. (a) Forward outcomes Y_N with forward protocol $\Lambda_N(Y_N)$. (b) Backward outcomes $Y_N^\dagger$ with backward protocol $[\Lambda_N(Y_N)^\dagger]$.

If the feedback control in the forward processes is "perfect," the particle is expected to return to its initial state with unit probability in the backward processes. In such a case, γ takes the maximum value that equals the number of possible outcomes of Y_N. In fact, for the case of the Szilard engine, $\gamma = 2$ holds corresponding to $W = -k_B T \ln 2$ and $\Delta F = 0$ [67]. In contrast, without feedback control, γ reduces to 1 as

$$\gamma := \int P[Y_N^\dagger] dY_N^\dagger = 1, \tag{90}$$

which vindicates the original integral fluctuation theorem. Therefore, the measurements in the backward processes are used to characterized to the efficacy of feedback control in the forward processes.

We stress that σ and γ can be measured independently, because σ is obtained from the forward experiments with feedback, and γ is obtained from the backward experiments without feedback. Therefore, Eqs. (87) and (89) can be directly verified in experiments. In fact, Eq. (89) has been verified in a real experiment by using a feedback-controlled ratchet with a Brownian particle [82].

From Eq. (65), we have the second generalization of the second law of thermodynamics

$$\langle \sigma \rangle \geqslant -\ln \gamma. \tag{91}$$

The equality in inequality (91) is achieved if σ does not fluctuate. We note that, if the distribution of σ is Gaussian, we have a generalized fluctuation-dissipation theorem

$$\langle \sigma \rangle + \ln \gamma = \tfrac{1}{2}(\langle \sigma^2 \rangle - \langle \sigma \rangle^2). \tag{92}$$

While the first generalization (66) involves only the term of the obtained information, the second generalization (87) involves the term of feedback efficacy. To understand the relationship between the mutual information I_c and the feedback efficacy γ, we introduce the notation

$$C[A] := -\ln \langle e^{-A} \rangle \tag{93}$$

for any probability variable A. We note that, if A can be written as $A = tA'$ with t being a real number and A' being another probability variable, then $C[A]$ is the cumulant generation function of A'. By using this notation, we have

$$C[\sigma] + C[I_c] - C[\sigma + I_c] = -\ln \gamma, \tag{94}$$

because $C[\sigma] = -\ln \gamma$ in Eq. (87), $C[I_c] = 0$ holds as in Eq. (56), and $C[\sigma + I_c] = 0$ holds as in Eq. (66). Equality (94) implies that $-\ln \gamma$ is a measure of the correlation between σ and I_c. This can be more clearly seen by the cumulant expansion of Eq. (94) if the joint distribution of σ and I_c is Gaussian:

$$\langle \sigma I_c \rangle - \langle \sigma \rangle \langle I_c \rangle = -\ln \gamma. \tag{95}$$

Therefore, γ characterizes how efficiently we use the obtained information to decrease the entropy production by feedback control: If γ is large, the more I_c we obtain, the less σ is.

We can also derive another nonequilibrium equality, which also gives us the information about the feedback efficacy. By taking logarithm of the both sides of Eq. (65), we obtain

$$\langle \sigma' \rangle = \int dY_N P[Y_N] \ln \frac{P[Y_N]}{P[Y_N^\dagger | \Lambda_N(Y_{N-1})^\dagger]}, \tag{96}$$

which is a generalization of Eq. (16). The same result under a different situation has also been obtained in Refs. [30,31]. Equality (96) implies that the renormalized entropy production equals the Kullback-Leibler divergence-like quantity between the forward probability $P[Y_N]$ and the backward probability $P[Y_N^\dagger | \Lambda_N(Y_{N-1})^\dagger]$. In fact, without feedback control, the right-hand side of Eq. (96) reduces to the Kullback-Leibler divergence between $P[Y_N]$ and $P^\dagger[Y_N^\dagger]$, and therefore the both sides of Eq. (96) are positive, which is consistent with the second law of thermodynamics. On the contrary, in the presence of feedback control, the right-hand side is no longer the Kullback-Leibler divergence, because $P[Y_N^\dagger | \Lambda_N(Y_{N-1})^\dagger]$ is not a normalized probability distribution in terms of $Y_N^\dagger$. Therefore the both sides of (96) can be negative. Since $\langle \sigma' \rangle \leqslant \langle \sigma \rangle$, the entropy production $\langle \sigma \rangle$ is bounded from below by the right-hand side of Eq. (96):

$$\langle \sigma \rangle \geqslant \int dY_N P[Y_N] \ln \frac{P[Y_N]}{P[Y_N^\dagger | \Lambda_N(Y_{N-1})^\dagger]}. \tag{97}$$

Without feedback control, the right-hand side of (97) gives a positive bound, while, with feedback control, the right-hand side can give a negative bound. We note that, for a quantum generalization of the Szilard engine with multiparticles, essentially the same result as Eq. (96) has been obtained [104].

We note that special cases of our results in this section were obtained elsewhere. We have derived two types of the generalized Jarzynski equality for the cases with a single measurement in the presence of a single heat bath in Ref. [67]. In Ref. [71] the detailed fluctuation theorem and the Jarzynski equality were obtained for the cases with multiple measurements and feedback in the presence of a single heat bath. In Ref. [68] a generalized Jarzynski equality was also obtained for the Kalman filter and the optimal control. The results in this paper include all of the above results and generalize them to the cases of multiple heat baths and non-Markovian measurements.

We also note that the generalized Jarzynski equality (89) with a single measurement was experimentally verified by using a feedback-controlled ratchet with a colloidal particle [82]. Moreover, Eq. (89) has been generalized to quantum systems [72].

VI. EXAMPLES

We now discuss two examples that illustrate the essential features of our general results. We analytically discuss a gener-

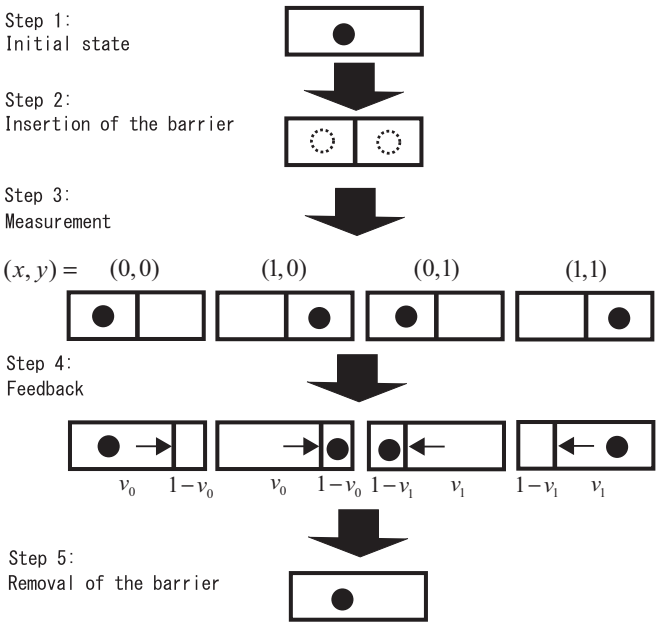

FIG. 4. Generalized Szilard engine with measurement error rate ε, where x denotes the states of the system, and y denote the measurement outcomes. The control protocol is determined by y.

alized Szilard engine with measurement errors in Sec. VI A and numerically discuss a feedback-controlled ratchet in Sec. VI B.

A. Szilard engine with measurement errors

As an example with a classical measurement, we discuss a generalized Szilard engine with measurement errors, which will be shown to achieve the upper bound of inequality (72) or (73) for an arbitrary error rate. The control protocol of the generalized Szilard engine is given by the following steps, which are described in Fig. 4.

Step 1. Initial state. A single-particle classical gas is in a box. The initial state of the gas is in thermal equilibrium with a single heat bath at temperature $T = (k_{\rm B}\beta)^{-1}$.

Step 2. Insertion of the barrier. We insert a barrier in the middle of the box and divide it into two boxes with the same volume. Here we do not know in which box the particle is. For simplicity of notations, we write "left" as "0" and "right" as "1." In other words, the position x of the particle is given by $x = 0$ or $x = 1$. We do not need any work during this process, as proved in Ref. [104].

Step 3. Measurement. We measure the position of the particle. We assume that the measurement is equivalent to the binary symmetric channel with error rate ε [106]; the measurement outcome takes $y = 0$ or 1, and the measurement error is characterized by conditional probabilities $P[0|0] = P[1|1] - 1 = \varepsilon$ and $P[0|1] = P[1|0] = \varepsilon$ with $0 \leqslant \varepsilon \leqslant 1$. We note that $x = y$ holds for the original Szilard engine without error ($\varepsilon = 0$).

Step 4. Feedback. We next move the position of the barrier quasistatically and isothermally. The protocol of moving the barrier depends on measurement outcome y. Let v_0 ($0 \leqslant v_0 \leqslant 1$) and v_1 ($0 \leqslant v_0 \leqslant 1$) be real numbers. We assume that, after we move the barrier, the ratio of the volumes of the boxes is assumed to be $v_0 : 1 - v_0$ for $y = 0$, or $1 - v_1 : v_1$ for $y = 1$. We note that, in the case of the original Szilard engine, $v_0 = v_1 = 1$ holds. In this process, we extract the work from the engine. The amounts of the work are given by $k_{\rm B}T \ln 2v_0$ if $(x,y) = (0,0)$, $k_{\rm B}T \ln 2(1 - v_0)$ if $(x,y) = (0,1)$, $k_{\rm B}T \ln 2(1 - v_1)$ if $(x,y) = (1,0)$, and $k_{\rm B}T \ln 2v_1$ if $(x,y) = (1,1)$. The feedback protocol is characterized by v_0 and v_1.

Step 5. Removal of the barrier. We remove the barrier without any work. The engine then returns to the initial state.

From the total process, we extract the average work

$$\langle W_{\rm ext}\rangle = k_{\rm B}T\Bigg[\ln 2 + \frac{1-\varepsilon}{2}\ln v_0 + \frac{\varepsilon}{2}\ln(1-v_0) + \frac{\varepsilon}{2}\ln(1-v_1) + \frac{1-\varepsilon}{2}\ln v_1\Bigg]. \tag{98}$$

We note that $\Delta F^{\rm S} = 0$ holds. We then maximize $\langle W_{\rm ext}\rangle$ under a given measurement error ε by changing v_0 and v_1. The maximum value of $\langle W_{\rm ext}\rangle$ is achieved when

$$v_0 = v_1 = 1 - \varepsilon, \tag{99}$$

for which the maximum work is given by

$$\langle W_{\rm ext}\rangle = k_{\rm B}T[\ln 2 + \varepsilon\ln\varepsilon + (1-\varepsilon)\ln(1-\varepsilon)]. \tag{100}$$

On the other hand, the mutual information of the binary symmetric channel is given by

$$\langle I\rangle = \ln 2 + \varepsilon\ln\varepsilon + (1-\varepsilon)\ln(1-\varepsilon). \tag{101}$$

Therefore, we obtain

$$\langle W_{\rm ext}\rangle = k_{\rm B}T\langle I\rangle, \tag{102}$$

which means that the generalized Szilard engine achieves the upper bound of the extractable work (72) or (73) for any amount of the mutual information.

We also check the generalized Jarzynski equalities in this model for arbitrary v_0, v_1, and ε. We first note that $I[x:y]$ is given by $\ln 2(1-\varepsilon)$ when $(x,y) = (0,0)$, $\ln 2\varepsilon$ when $(x,y) = (0,1)$, $\ln 2\varepsilon$ when $(x,y) = (1,0)$, and $\ln 2(1-\varepsilon)$ when $(x,y) = (1,1)$. Therefore we obtain

$$\langle e^{-\beta W - I}\rangle = \frac{v_0 + (1-v_0) + (1-v_1) + v_1}{2} = 1, \tag{103}$$

which confirms Eq. (71).

We next consider the second generalization (89) of the Jarzynski equality. Corresponding to two measurement outcomes $y = 0,1$, we have two backward control protocols as follows (see also Fig. 5).

Step 1. Initial state. The initial state of the backward control is in the thermal equilibrium.

Step 2. Insertion of the barrier. Corresponding to Step 5 of the forward process, we insert the barrier and divide the box into two boxes, because the time reversal of the barrier removal is the barrier insertion. Corresponding to $y = 0$ or $y = 1$ in the forward process, we divide the box with the ratio $v_0 : 1 - v_0$ or $1 - v_1 : v_1$, respectively.

Step 3. Moving the barrier. We next move the barrier to the middle of the box quasistatically and isothermally. This is the time reversal of the feedback control in Step 4 of the forward process.

Step 4. Measurement. We perform the measurement to find in which box the particle is in. Corresponding to the backward

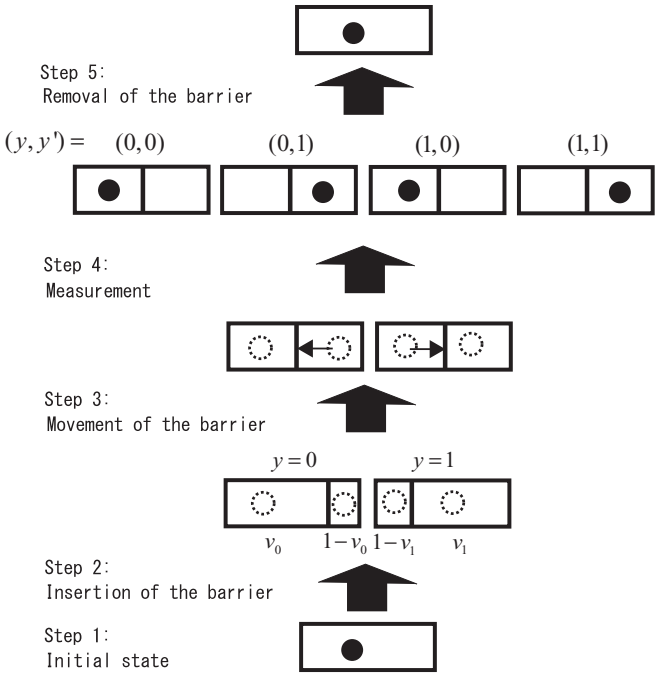

FIG. 5. Backward processes of the generalized Szilard engine. Corresponding to y that denotes the measurement outcomes in the forward process, we have two control protocols in the backward process, where y' denotes the measurement outcomes in the backward process.

protocol with $y = 0$, we obtain the outcomes of backward measurement $y' = 0$ with probability $P[y' = 0|\Lambda(y = 0)^\dagger] = v_0(1-\varepsilon) + (1-v_0)\varepsilon$ and $y' = 1$ with probability $P[y' = 1|\Lambda(y = 0)^\dagger] = v_0\varepsilon + (1-v_0)(1-\varepsilon)$. On the other hand, corresponding to the backward protocol with $y = 1$, we obtain the outcomes of backward measurement $y' = 0$ with probability $P[y' = 0|\Lambda(y = 1)^\dagger] = v_1\varepsilon + (1-v_1)(1-\varepsilon)$ and $y' = 1$ with probability $P[y' = 1|\Lambda(y = 1)^\dagger] = v_1(1-\varepsilon) + (1-v_0)\varepsilon$.

Step 5. Removal of the barrier. We remove the barrier, and the system returns to the initial state. This is the time reversal of the barrier insertion in Step 2 of the forward process.

From Step 4 of the backward process, we have

$$\gamma := P[y' = 0|\Lambda(y = 0)^\dagger] + P[y' = 1|\Lambda(y = 1)^\dagger] = (1-\varepsilon)(v_0 + v_1) + \varepsilon(2 - v_0 - v_1). \quad (104)$$

On the other hand, we can straightforwardly obtain

$$\langle e^{-\beta W}\rangle = (1-\varepsilon)(v_0 + v_1) + \varepsilon(2 - v_0 - v_1), \quad (105)$$

which confirms Eq. (89).

B. Feedback-controlled ratchet

We next discuss a model for Brownian motors [110–115], in particular a feedback-controlled ratchet [58,61,63]. We consider a rotating Brownian particle with a periodic boundary condition. Let x be the position or the angle of the particle, and its boundary condition is given by $x = x + L$ with L being a constant. In the following, we restrict the particle's position to $-L/2 \leqslant x < L/2$. We assume that the particle obeys the overdamped Langevin equation (32) and that control

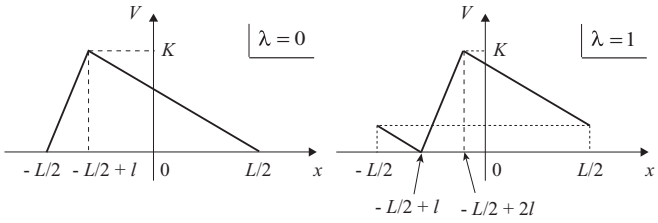

FIG. 6. Two shapes of potential $V(x,\lambda)$ corresponding to $\lambda = 0,1$.

parameter λ takes two values ($\lambda = 0$ or 1). Corresponding to them, the ratchet potential V takes the following two profiles (Fig. 6):

$$V(x,0) = \begin{cases} K(x + L/2)/l \\ \quad (-L/2 \leqslant x < -L/2 + l), \\ -K(x - L/2)/(L - l) \\ \quad (-L/2 + l \leqslant x < L), \end{cases} \quad (106)$$

$$V(x,1) = \begin{cases} -K(x + L/2 - l)/(L - l) \\ \quad (-L/2 \leqslant x < -L/2 + l), \\ K(x + L/2 + 2l)/l \\ \quad (-L/2 + l \leqslant x < -L/2 + 2l), \\ -K(x - L/2 - l)/(L - l) \\ \quad (-L/2 + 2l \leqslant x < L/2), \end{cases} \quad (107)$$

where l is a constant with $0 < l < L/2$, and K is a positive constant that characterizes the height of the potential.

We start with the initial equilibrium with parameter $\lambda = 0$ and control the system from time $t = 0$ to τ with the following three protocols:

(1) *Trivial control.* We do not change the parameter $\lambda = 0$.

(2) *Flashing ratchet.* At times $t = m\tau_0$ with m being integers and τ_0 being a constant, we switch parameter λ from 0 to 1 or from 1 to 0 periodically.

(3) *Feedback-controlled ratchet.* At times $t = m\tau_0$, we switch the parameter with the following feedback protocol. We measure the position x at $t = m\tau_0$ without error. We then set $\lambda = 1$ from $t = m\tau_0$ to $(m+1)\tau_0$ if and only if the outcome is in $-L/2 \leqslant x < -L/2 + l$. Otherwise, parameter λ is set to 0.

For numerical simulations, we set $l = 3L/10$, $K = 3k_BT$, $\tau_0 = 0.05$, and $\tau = 0.25$, with units $k_BT = 1$, $L = 1$, and $\eta/2 = 1$. We performed the simulations by discretizing Eq. (32) with $\Delta t = 0.00025$ for 1 000 000 samples. We note that, to obtain the initial thermal equilibrium, we waited $\tau_{\rm wait} = 0.5$ and checked that the system was fully thermalized in the periodic ratchet with parameter $\lambda = 0$.

The time evolution of the ensemble average $\langle x(t)\rangle$ is plotted in Fig. 7(a) for the above three protocols. As expected, nothing happens for the first protocol, while the particle is transported to the right on average for the second and third protocols. In the case of the feedback-controlled ratchet, the particle is transported to the right faster than the case of the flashing ratchet. Figure 7(b) shows the time evolution of the work $\langle W(t)\rangle$ that is performed on the particle. The work is induced only in the switching times. We find that, in order to transport the particle, the energy input to the particle with feedback control is smaller than that with the flashing.

Figure 8 shows the left-hand side of the Jarzynski equality $\langle e^{-\beta(W-\Delta F)}\rangle$ for the flashing and feedback-controlled ratchet,

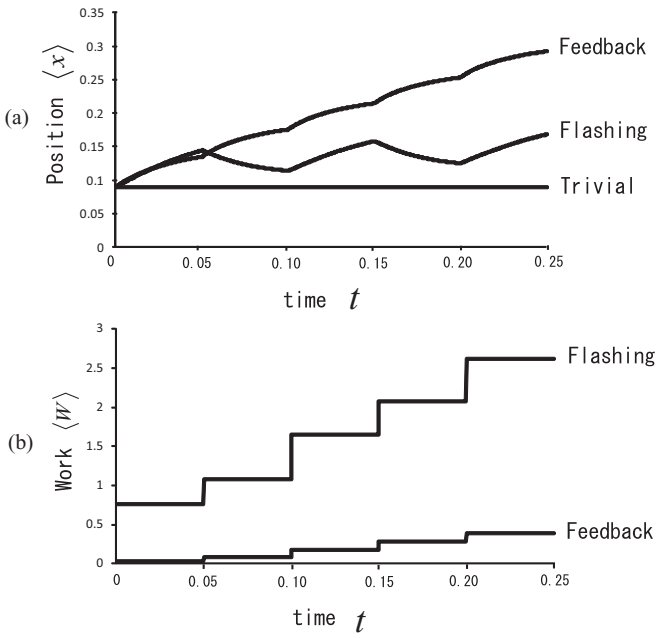

FIG. 7. (a) Numerical results of the ensemble average of trajectory $x(t)$ corresponding to the three control protocols: the trivial control, the flashing ratchet, and the feedback-controlled ratchet. (b) Numerical result of the ensemble average of the work $W(t)$ corresponding to the flashing ratchet and the feedback-controlled ratchet.

and the efficacy parameter γ for the feedback-controlled ratchet. We note that $\Delta F = 0$ always holds. With feedback control, $\langle e^{-\beta(W-\Delta F)}\rangle$ increases from 1 as the number of switchings increases, while, without feedback control, $\langle e^{-\beta(W-\Delta F)}\rangle$ converges to 1 for all switching times in consistent with the original Jarzynski equality. On the other hand, to obtain γ, we numerically performed the backward experiments. The discretization of the time is $\Delta t = 0.0005$, and the number of the samples is 10 000 for each trajectory of $\lambda(t)$. We note that the number of the trajectories of λ is given by 2^m with m times of switchings. Figure 8 shows a good coincidence between $\langle e^{-\beta W}\rangle$ and γ, which confirms the validity of Eq. (89) in the feedback-controlled ratchet.

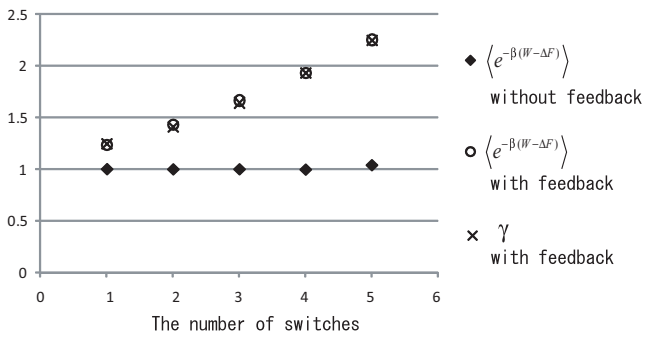

FIG. 8. Numerical tests of the Jarzynski equality for the flashing ratchet and a generalized Jarzynski equality (89) for the feedback-controlled ratchet.

VII. CONCLUSIONS

In this paper we have studied the effects of measurements and feedback control on nonequilibrium thermodynamic systems. In particular, we have generalized nonequilibrium equalities to the systems that are subject to feedback control. Our formulations and results are applicable to a broad class of classical nonequilibrium systems.

In Sec. II we reviewed stochastic thermodynamics by focusing on the nonequilibrium equalities. In Sec. III we formulated measurements on nonequilibrium systems and defined mutual information I_c by Eq. (31) for multiple measurements. In Sec. IV we formulated feedback control on nonequilibrium systems. We discussed the properties of the joint probability (47), which is well defined due to causality. We introduced the mutual information I_c by Eq. (55), which is not equivalent to I in the presence of feedback control. In fact, I_c describes the correlation between the system and the outcomes, which characterizes the effective information obtained by the measurements. We have also shown that the detailed fluctuation theorem (58) holds in the presence of feedback control.

Section V constitutes the main results of this paper. We derived two types of generalizations of the nonequilibrium equalities. In Sec. V A we derived a generalized detailed fluctuation theorem (65), which involves the mutual information. Based on Eq. (65), we derived the generalizations of the integral fluctuation theorem (66), the Jarzynski equality (71), the second laws [(67), (72), and (73)], the fluctuation-dissipation theorem (70), and the KPB equality (68), which all involve the mutual information. In Sec. V B, we derived the renormalized detailed fluctuation theorem (83) and derived the generalizations of the integral fluctuation theorem (87), the Jarzynski equality (89), the second law (91), the fluctuation-dissipation theorem (92), and the KPB equality (96). We have shown that mutual information I_c, rather than I, plays the crucial role to formulate the nonequilibrium equalities under feedback control. These results are the generalizations of the fundamental equalities in nonequilibrium statistical mechanics to feedback-controlled processes, and lead to the generalized second law of thermodynamics with feedback control, which gives the minimal energy cost that is needed for the feedback control.

In Sec. VI we discussed simple examples to explicitly show that our results in Sec. V can be applied to typical situations. In Sec. VI A, we discussed the Szilard engine with measurement errors that achieves the equality of the generalized second law of thermodynamics (72) or (73). This is an important model to quantitatively illustrate that the mutual information can be converted to the work. We also confirmed the two generalized Jarzynski equalities (71) and (89) in the generalized Szilard engine. In Sec. VI B, we considered a feedback-controlled ratchet and confirmed a generalized Jarzynski equality (89).

All of our formulations and results are consistent with the original nonequilibrium equalities and the second law of thermodynamics, and our results serve as the fundamental principle of nonequilibrium thermodynamics of feedback control. We note that, in our results such as Eq. (65), the thermodynamic quantities and the information contents are treated on an equal footing. Therefore, our theory may

be regarded as the nonequilibrium version of "information thermodynamics" [100,102], which serves as the fundamental theory of nonequilibrium information heat engines.

ACKNOWLEDGMENTS

We are grateful to Y. Fujitani, H. Hayakawa, H. Hasegawa, J. M. Horowitz, S. Ito, K. Kawaguchi, T. S. Komatsu, N. Nakagawa, K. Saito, M. Sano, S. Sasa, H. Suzuki, H. Tasaki, and S. Toyabe for valuable discussions. This work was supported by a Grant-in Aid for Scientific Research on Innovative Areas "Topological Quantum Phenomena" (KAKENHI 22103005) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, and by a Global COE program "Physical Science Frontier" of MEXT, Japan. TS acknowledges JSPS Research Fellowships for Young Scientists (Grant No. 208038) and the Grant-in-Aid for Research Activity Start-up (Grant No. 11025807).

APPENDIX: PHYSICAL MEANING OF THE ENTROPY PRODUCTION

In this Appendix, we discuss the physical meanings of the entropy production σ in the following two typical setups to clarify the typical situations to which our results apply.

Isothermal processes. We assume that there is a single heat bath at temperature $T = (k_B\beta)^{-1}$ and that the initial distributions of both forward and backward experiments are in the canonical distributions. We stress that we do not assume that the final distributions of both the forward and backward experiments are in the canonical distributions: The final distribution of the forward (backward) experiments does not necessarily equal the initial distribution of the backward (forward) experiments. Let $H(x,\lambda)$ be the Hamiltonian of the system with the time symmetry $H(x,\lambda) = H(x^*,\lambda^*)$. The canonical distribution with parameter λ is given by

$$P_{\rm can}[x|\lambda] := e^{\beta[F(\lambda)-H(x,\lambda)]}, \tag{A1}$$

where

$$F(\lambda) := -k_B T \ln \int dx e^{-\beta H(x,\lambda)} \tag{A2}$$

is the Helmholtz free energy. In this situation, the entropy production reduces to

$$\sigma[X_N] = \beta(W[X_N] - \Delta F), \tag{A3}$$

where

$$W[X_N] := H(x_N,\lambda_{\rm fin}) - H(x_0,\lambda_{\rm int}) - Q[X_N] \tag{A4}$$

is the work performed on the system from the external parameter, and $\Delta F := F(\lambda_{\rm fin}) - F(\lambda_{\rm int})$ is the free-energy difference. In this case Eq. (14) leads to the Jarzynski equality (19), and the second law (15) reduces to inequality (20).

Transition between arbitrary nonequilibrium states. We assume that there are several heat baths, and that we can control the strength of interaction between the system and the baths through λ. In other words, we can attach or detach the system from the baths by controlling λ; for example, we can attach an adiabatic wall to the system. We set an arbitrary initial distribution $P_0[x_0]$ for the forward experiments. On the other hand, the initial state of the backward experiments is assumed to be taken as $P_0^\dagger[x_0^\dagger] := P_N[x_N]$, where $P_N[x_N]$ is the final distribution of the forward experiments. Although this choice of the backward initial state is artificial and is difficult to be experimentally realized except for special cases, this backward initial state is a theoretically useful tool to derive a version of the second law of thermodynamics as follows. In this case the entropy production is given by

$$\sigma[X_N] = -\ln P_N[x_N] + \ln P_0[x_0] - \sum_i \beta_i Q_i[X_N], \tag{A5}$$

and its ensemble average leads to

$$\langle\sigma\rangle = S_N - S_0 - \sum_i \beta_i \langle Q_i\rangle, \tag{A6}$$

where

$$S_n := -\int P_n[x_n] \ln P_n[x_n] dx_n \tag{A7}$$

is the Shannon entropy at time t_n. By introducing notation $\Delta S := S_N - S_0$, the second law (15) leads to

$$\Delta S \geqslant \sum_i \beta_i \langle Q_i\rangle. \tag{A8}$$

[1] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum, *Feedback Control Theory* (Macmillan, New York, 1992).
[2] K. J. Åstrom and R. M. Murray, *Feedback Systems: An Introduction for Scientists and Engineers* (Princeton University Press, Princeton, 2008).
[3] K. Sekimoto, Prog. Theor. Phys. Suppl. **130**, 17 (1998).
[4] C. Bustamante, J. Liphardt, and F. Ritort, Phys. Today **58**, 43 (2005).
[5] U. Seifert, Euro. Phys. J. B **64**, 423 (2008).
[6] K. Sekimoto, *Stochastic Energetics* (Springer-Verlag, Berlin, 2010).
[7] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993).
[8] G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995).
[9] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[10] J. Kurchan, J. Phys. A. Math. Gen. **31**, 3719 (1998).
[11] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).
[12] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).
[13] J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).
[14] C. Maes, J. Stat. Phys. **95**, 367 (1999).
[15] C. Maes, F. Redig, and A. Van Moffaert, J. Math. Phys. **41**, 1528 (2000).

[16] C. Jarzynski, J. Stat. Phys. **98**, 77 (2000).
[17] J. Kurchan, e-print arXiv:cond-mat/0007360.
[18] H. Tasaki, e-print arXiv:cond-mat/0009244.
[19] T. Hatano and S.-I. Sasa, Phys. Rev. Lett. **86**, 3463 (2001).
[20] D. J. Evans and D. J. Searles, Adv. Phys. **51**, 1529 (2002).
[21] R. van Zon and E. G. D. Cohen, Phys. Rev. Lett. **91**, 110601 (2003).
[22] C. Jarzynski, J. Stat. Mech. (2004) P09005.
[23] C. Jarzynski and D. K. Wójcik, Phys. Rev. Lett. **92**, 230602 (2004).
[24] T. Harada and S.-I. Sasa, Phys. Rev. Lett. **95**, 130602 (2005).
[25] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).
[26] M. Esposito and S. Mukamel, Phys. Rev. E **73**, 046129 (2006).
[27] D. Andrieux and P. Gaspard, J. Stat. Mech. (2007) P02006.
[28] K. Saito and A. Dhar, Phys. Rev. Lett. **99**, 180601 (2007).
[29] T. Ohkuma and T. Ohta, J. Stat. Mech. (2007) P10010.
[30] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. Lett. **98**, 080602 (2007).
[31] A. Gomez-Marin, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. E **78**, 011107 (2008).
[32] T. S. Komatsu and N. Nakagawa, Phys. Rev. Lett. **100**, 030601 (2008).
[33] T. S. Komatsu, N. Nakagawa, S. I. Sasa, and H. Tasaki, Phys. Rev. Lett. **100**, 230602 (2008).
[34] Y. Utsumi and K. Saito, Phys. Rev. B **79**, 235311 (2009).
[35] M. Campisi, P. Talkner, and P. Hänggi, Phys. Rev. Lett. **102**, 210401 (2009).
[36] J. Ren, P. Hänggi, and B. Li, Phys. Rev. Lett. **104**, 170601 (2010).
[37] H. Hasegawa, J. Ishikawa, K. Takara, and D. J. Driebe, Phys. Lett. A **374**, 1001 (2010).
[38] M. Esposito and C. Van den Broeck, Phys. Rev. Lett. **104**, 090601 (2010).
[39] M. Campisi, P. Talkner, and P. Hänggi, Phys. Rev. Lett. **105**, 140601 (2010).
[40] S. Vaikuntanathan and C. Jarzynski, Euro. Phys. Lett. **87**, 60005 (2010).
[41] M. Campisi, P. Hänggi, and P. Talkner, Rev. Mod. Phys. **83**, 771 (2011).
[42] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **89**, 050601 (2002).
[43] J. Liphardt *et al.*, Science **296**, 1832 (2002).
[44] E. H. Trepagnier *et al.*, Proc. Natl. Acad. Sci. USA **101**, 15038 (2004).
[45] D. M. Carberry, J. C. Reid, G. M. Wang, E. M . Sevick, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **92**, 140601 (2004).
[46] D. Collin *et al.*, Nature (London) **437**, 231 (2005).
[47] F. Douarche, S. Joubaud, N. B. Garnier, A. Petrosyan, and S. Ciliberto, Phys. Rev. Lett. **97**, 140603 (2006).
[48] D. Andrieux, P. Gaspard, S. Ciliberto, N. Garnier, S. Joubaud, and A. Petrosyan, Phys. Rev. Lett. **98**, 150601 (2007).
[49] S. Toyabe, H. R. Jiang, T. Nakamura, Y. Murayama, and M. Sano, Phys. Rev. E **75**, 011122 (2007).
[50] S. Toyabe, T. Okamoto, T. Watanabe-Nakayama, H. Taketani, S. Kudo, and E. Muneyuki, Phys. Rev. Lett. **104**, 198103 (2010).
[51] K. Hayashi, H. Ueno, R. Iino, and H. Noji, Phys. Rev. Lett. **104**, 218103 (2010).
[52] S. Nakamura *et al.*, Phys. Rev. Lett. **104**, 080602 (2010).
[53] V. Serreli *et al.*, Nature (London) **445**, 523 (2007).
[54] S. Rahav, J. Horowitz, and C. Jarzynski, Phys. Rev. Lett. **101**, 140602 (2008).
[55] E. R. Kay, D. A. Leigh, and F. Zerbetto, Angew. Chem. **46**, 72 (2007).
[56] H. Gu *et al.*, Nature (London) **465**, 202 (2010).
[57] M. Schliwa and G. Woehlke, Nature (London) **422**, 759 (2003).
[58] F. J. Cao, L. Dinis, and J. M. R. Parrondo, Phys. Rev. Lett. **93**, 040603 (2004).
[59] K. H. Kim and H. Qian, Phys. Rev. Lett. **93**, 120602 (2004).
[60] K. H. Kim and H. Qian, Phys. Rev. E **75**, 022102 (2007).
[61] B. J. Lopez, N. J. Kuwada, E. M. Craig, B. R. Long, and H. Linke, Phys. Rev. Lett. **101**, 220601 (2008).
[62] F. J. Cao and M. Feito, Phys. Rev. E **79**, 041118 (2009).
[63] M. Feito, J. P. Baltanas, and F. J. Cao, Phys. Rev. E **80**, 031128 (2009).
[64] F. J. Cao, M. Feito, and H. Touchette, Physica A **388**, 113 (2009).
[65] M. Bonaldi *et al.*, Phys. Rev. Lett. **103**, 010601 (2009).
[66] H. Suzuki and Y. Fujitani, J. Phys. Soc. Jpn. **78**, 074007 (2009).
[67] T. Sagawa and M. Ueda, Phys. Rev. Lett. **104**, 090602 (2010).
[68] Y. Fujitani and H. Suzuki, J. Phys. Soc. Jpn. **79**, 104003 (2010).
[69] T. Brandes, Phys. Rev. Lett. **105**, 060602 (2010).
[70] M. Ponmurugan, Phys. Rev. E **82**, 031129 (2010).
[71] J. M. Horowitz and S. Vaikuntanathan, Phys. Rev. E **82**, 061120 (2010).
[72] Y. Morikuni and H. Tasaki, J. Stat. Phys. **13**, 1 (2011).
[73] S. Ito and M. Sano, Phys. Rev. E **84**, 021123 (2011).
[74] D. Abreu and U. Seifert, Euro. Phys. Lett. **94**, 10001 (2011).
[75] J. M. Horowitz and J. M. R. Parrondo, Euro. Phys. Lett. **95**, 10005 (2011).
[76] M. Esposito and C. Van den Broeck, Euro. Phys. Lett. **95**, 40004 (2011).
[77] S. Vaikuntanathan and C. Jarzynski, Phys. Rev. E **83**, 061120 (2011).
[78] T. Sagawa, J. Phys.: Conf. Ser. **297**, 012015 (2011).
[79] S. Lahiri, S. Rana, and A. M. Jayannavar, J. Phys. A: Math. Theor. **45**, 065002 (2012).
[80] D. V. Averin, M. Möttönen, and J. P. Pekola, Phys. Rev. B **84**, 245448 (2011).
[81] J. M. Horowitz and J. M. R. Parrondo, New J. Phys. **13**, 123019 (2011).
[82] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano, Nat. Phys. **6**, 988 (2010).
[83] J. C. Maxwell, *Theory of Heat* (Appleton, London, 1871).
[84] H. S. Leff and A. F. Rex, eds., *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing* (Princeton University Press, Princeton, 2003).
[85] L. Szilard, Z. Phys. **53**, 840 (1929).
[86] L. Brillouin, J. Appl. Phys. **22**, 334 (1951).
[87] C. H. Bennett, Int. J. Theor. Phys. **21**, 905 (1982).
[88] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961).
[89] S. Lloyd, Phys. Rev. A **39**, 5378 (1989).
[90] S. Lloyd and W. H. Zurek, J. Stat. Phys. **62**, 819 (1991).
[91] S. Lloyd, Phys. Rev. A **56**, 3374 (1997).
[92] H. Touchette and S. Lloyd, Phys. Rev. Lett. **84**, 1156 (2000).
[93] W. H. Zurek, e-print arXiv:quant-ph/0301076.
[94] M. O. Scully *et al.*, Science **299**, 862 (2003).
[95] H. Touchette and S. Lloyd, Physica A **331**, 140 (2004).
[96] T. D. Kieu, Phys. Rev. Lett. **93**, 140403 (2004).
[97] A. E. Allahverdyan *et al.*, J. Mod. Opt. **51**, 2703 (2004).
[98] H. T. Quan, Y. D. Wang, Y. X. Liu, C. P. Sun, and F. Nori, Phys. Rev. Lett. **97**, 180402 (2006).

[99] M. A. Nielsen, C. M. Caves, B. Schumacher, and H. Barnum, Proc. R. Soc. London A **454**, 277 (1998).
[100] T. Sagawa and M. Ueda, Phys. Rev. Lett. **100**, 080403 (2008).
[101] K. Jacobs, Phys. Rev. A **80**, 012322 (2009).
[102] T. Sagawa and M. Ueda, Phys. Rev. Lett. **102**, 250602 (2009); **106**, 189901(E) (2011).
[103] K. Maruyama, F. Nori, and V. Vedral, Rev. Mod. Phys. **81**, 1 (2009).
[104] S. W. Kim, T. Sagawa, S. De Liberato, and M. Ueda, Phys. Rev. Lett. **106**, 070401 (2011).
[105] C. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948).
[106] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley and Sons, New York, 1991).
[107] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[108] G. Welch and G. Bishop, Technical Report TR 95-041, Department of Computer Science, University of North Carolina (unpublished).
[109] D. P. Bertsekas, *Dynamic Programming and Optimal Control* (Athena Scientific, Belmont, 2005).
[110] R. D. Vale and F. Oosawa, Adv. Biophys. **26**, 97 (1990).
[111] F. Julicher, A. Ajdari, and J. Prost, Rev. Mod. Phys. **69**, 1269 (1997).
[112] J. M. R. Parrondo and B. J. De Cisneros, Appl. Phys. A **75**, 179 (2002).
[113] P. Reimann, Phys. Rep. **361**, 57 (2002).
[114] P. Hänggi, F. Marchesoni, and F. Nori, Ann. Phys. **14**, 51 (2005).
[115] P. Hänggi and F. Marchesoni, Rev. Mod. Phys. **81**, 387 (2009).