Fluctuation theorem for hidden entropy production

Kyogo Kawaguchi* and Yohei Nakayama†
Department of Physics, The University of Tokyo, Hongo 7-3-1, Tokyo 113-0033, Japan
(Received 10 October 2012; published 29 August 2013)

Elimination of seemingly unnecessary variables in Markovian models may cause a difference in the value of irreversible entropy production between the original and reduced dynamics. We show that such difference, which we call the hidden entropy production, obeys an integral fluctuation theorem if all variables are time-reversal invariant, or if the density function is symmetric with respect to the change of sign of the time-reversal antisymmetric variables. The theorem has wide applicability, since the proposed condition is mostly satisfied in the case where the hidden fast variables are equilibrated. The main consequence of this theorem is that the entropy production decreases by the coarse-graining procedure. By contrast, in the case where a stochastic process is obtained by coarse-graining a deterministic and reversible dynamics, the entropy production may increase, implying that the integral fluctuation theorem should not hold for such reductions. We reveal, with an explicit example, that the nonequilibrated time-reversal antisymmetric variables play a crucial role in distinguishing these two cases, thus guaranteeing the consistency of the presented theorem.

DOI: 10.1103/PhysRevE.88.022147 PACS number(s): 05.20.−y, 05.70.−a

I. INTRODUCTION

Understanding and controlling micron- to nanoscale thermodynamic systems has been one of the central achievements in recent nonequilibrium statistical physics [1]. With the motivation to explore the design principles of biological molecular motors [2], or to construct engines that utilize thermal forces [3,4], particular interest has been devoted to the energetics and thermodynamic efficiencies in such heat-fluctuation dominated world. Though novel techniques have enabled us to precisely measure and manipulate the motion of Brownian particles [5], we have not yet been able to directly obtain quantities such as work, heat, and entropy production in those single molecular experiments. Nevertheless, owing to recent theoretical frameworks that allow concepts of thermodynamics to be applied to stochastic dynamics [6], we are capable of indirectly measuring such quantities by estimating the potential energy and external forces from the trajectories and the steady state distribution of the particle's position.

The fundamental Markovian description of the Brownian motion, the underdamped Langevin equation, takes into account both the velocity and position of the particle of interest. It is well known, however, that in realistic situations where colloidal particles are suspended in water, the velocity variable may be eliminated, and the position variable suffices to describe the Markovian dynamics (the overdamped Langevin description) [7]. This is a typical example where the same physical system may be modeled by two or more different levels of description. In the study of statistical mechanics, or more generally in the process of model analysis, the reduced and simplified descriptions are often related to the more detailed and fundamental equations through a mathematical procedure, for instance, by considering the time scale separation [8]. It is therefore naturally assumed in many cases that despite the arbitrariness in choosing the set of variables, the basic physics remains the same in all the levels of description.

However, in the case where a Brownian particle is put in an inhomogeneous temperature field, it has been found [9,10] that the underdamped and overdamped descriptions provide different values of the irreversible entropy production. We call such difference in the values of entropy production, the hidden entropy production, since it is an invisible quantity from the point of view of the reduced dynamics. As described above, the measurement of thermodynamic quantities in small systems is performed indirectly based on the notion that the particle is undergoing overdamped dynamics. The presence of hidden entropy production therefore causes a problem in defining and measuring the thermodynamic efficiency of a Brownian heat engine [11,12], and more generally in considering the physical meaning of the entropy production in each level of the model descriptions.

In this article, we provide a general theory on this hidden entropy production, which is directly related to whether the entropy production increases or decreases by the reduction of dynamics. First, we present an integral fluctuation theorem [Eq. (11)] obeyed by the hidden entropy production. From this theorem we may prove that the entropy production decreases on average by the reduction of dynamics. Second, by clarifying the condition for Eq. (11) to hold, we point out that the entropy production is allowed to increase on average by coarse-graining only when the density function for time-reversal antisymmetric variables is asymmetric in the original system. We show that this is indeed the case when the extended multibaker model is reduced to a simple random walk.

II. MODEL AND DEFINITIONS

We consider the Markov chain dynamics on a continuous state space. The continuous variables x and y may each represent many variables, nevertheless we use a single variable notation. The time evolution of the probability density function

*kyogok@daisy.phys.s.u-tokyo.ac.jp
†nakayama@daisy.phys.s.u-tokyo.ac.jp

$P_t(x,y)$ follows the equation

$$P_{t+\Delta t}(x,y) = \int dx'\, dy'\, P_t(x',y') W_{\lambda(t)}(x,y|x',y'), \quad (1)$$

where Δt is the (infinitesimal) time step. $W_{\lambda(t)}(x,y|x',y')$ is the transition probability from (x',y') to (x,y) between time t and $t+\Delta t$, and we assume that it is controlled by the time-dependent external parameter $\lambda(t)$. The integral by x' and y' in Eq. (1) is taken over the whole space, and we have set $\int dx\, dy W_{\lambda(t)}(x,y|x',y') = 1$.

Let $\boldsymbol{x}_N = (x_0,x_1,\ldots,x_N)$, $\boldsymbol{y}_N = (y_0,y_1,\ldots,y_N)$ be the stochastic path taken by (x,y) during the N time steps starting from $t=0$. We define the stochastic Shannon entropy difference assigned to this path as

$$s(\boldsymbol{x}_N,\boldsymbol{y}_N) := \log \frac{P_0(x_0,y_0)}{P_{N\Delta t}(x_N,y_N)}. \quad (2)$$

Defining the path transition probability as $W_\lambda(\boldsymbol{x}_N,\boldsymbol{y}_N|x_0,y_0) := \prod_{i=0}^{N-1} W_{\lambda(i\Delta t)}(x_{i+1},y_{i+1}|x_i,y_i)$, the path probability as $P_\lambda(\boldsymbol{x}_N,\boldsymbol{y}_N) := P_0(x_0,y_0)W_\lambda(\boldsymbol{x}_N,\boldsymbol{y}_N|x_0,y_0)$, and the Shannon entropy of the whole system at time t as $S(t) := -\int dx\, dy P_t(x,y)\log P_t(x,y)$, we have $\langle s(\boldsymbol{x}_N,\boldsymbol{y}_N)\rangle_{\lambda,N} = S(N\Delta t) - S(0)$. Here the bracket $\langle\cdot\rangle_{\lambda,N}$ denotes the average $\int d\boldsymbol{x}_N\, d\boldsymbol{y}_N P_\lambda(\boldsymbol{x}_N,\boldsymbol{y}_N)\cdot$, where $d\boldsymbol{x}_N := \prod_{i=0}^N dx_i$ and $d\boldsymbol{y}_N := \prod_{i=0}^N dy_i$. Using $\widetilde{P}_t(x) := \int dy\, P_t(x,y)$, we further define the coarse-grained stochastic Shannon entropy difference as

$$\widetilde{s}(\boldsymbol{x}_N) := \log \frac{\widetilde{P}_0(x_0)}{\widetilde{P}_{N\Delta t}(x_N)}. \quad (3)$$

The Boltzmann constant is set to 1 throughout the paper.

Next we define the heat entropy production. The reverse trajectory of $(\boldsymbol{x}_N,\boldsymbol{y}_N)$ is written as $(\boldsymbol{x}_N^\dagger,\boldsymbol{y}_N^\dagger)$, where $\boldsymbol{x}_N^\dagger = (\bar{x}_N,\bar{x}_{N-1},\ldots,\bar{x}_0)$, $\boldsymbol{y}_N^\dagger = (\bar{y}_N,\bar{y}_{N-1},\ldots,\bar{y}_0)$ with $\bar{x}_i$ being the time reversal of x_i. Now the heat entropy production corresponding to the N step trajectory is

$$\sigma(\boldsymbol{x}_N,\boldsymbol{y}_N) := \log \frac{W_\lambda(\boldsymbol{x}_N,\boldsymbol{y}_N|x_0,y_0)}{W_{\lambda^\dagger}(\boldsymbol{x}_N^\dagger,\boldsymbol{y}_N^\dagger|\bar{x}_N,\bar{y}_N)}. \quad (4)$$

Here, $W_{\lambda^\dagger}$ is the transition probability assigned to the time-reversed protocol $W_{\lambda^\dagger}(\boldsymbol{x}_N^\dagger,\boldsymbol{y}_N^\dagger|\bar{x}_N,\bar{y}_N) := \prod_{i=0}^{N-1} W_{\bar{\lambda}[(N-i)\Delta t]}(\bar{x}_i,\bar{y}_i|\bar{x}_{i+1},\bar{y}_{i+1})$, which is defined using $\bar{\lambda}(t)$, the time reversal of $\lambda(t)$ [for example, if the control is by the magnetic field, $\bar{\lambda}(t)$ corresponds to $\lambda(t)$ with reversed direction]. It is known that $\sigma(\boldsymbol{x}_N,\boldsymbol{y}_N)$ corresponds to the entropy production induced by the energy transfer from the variables (x,y) to the hidden degrees of freedom in general stochastic models [13,14], and in the Hamiltonian system including heat baths [15]. Let the coarse-grained path transition probabilities be

$$\widetilde{W}_\lambda[\boldsymbol{x}_N|x_0,P_0(\cdot)] := \int d\boldsymbol{y}_N \frac{P_\lambda(\boldsymbol{x}_N,\boldsymbol{y}_N)}{\widetilde{P}_0(x_0)}, \quad (5)$$

$$\widetilde{W}_{\lambda^\dagger}[\boldsymbol{x}_N^\dagger|\bar{x}_N,P_{N\Delta t}(\cdot)] := \int d\boldsymbol{y}_N^\dagger \frac{P_{\lambda^\dagger}(\boldsymbol{x}_N^\dagger,\boldsymbol{y}_N^\dagger)}{\widetilde{P}_{N\Delta t}(\bar{x}_N)}, \quad (6)$$

where $P_{\lambda^\dagger}(\boldsymbol{x}_N^\dagger,\boldsymbol{y}_N^\dagger) := P_{N\Delta t}(\bar{x}_N,\bar{y}_N)\, W_{\lambda^\dagger}(\boldsymbol{x}_N^\dagger,\boldsymbol{y}_N^\dagger|\bar{x}_N,\bar{y}_N)$. Then the coarse-grained heat entropy production is defined as

$$\widetilde{\sigma}(\boldsymbol{x}_N) := \log \frac{\widetilde{W}_\lambda[\boldsymbol{x}_N|x_0,P_0(\cdot)]}{\widetilde{W}_{\lambda^\dagger}[\boldsymbol{x}_N^\dagger|\bar{x}_N,P_{N\Delta t}(\cdot)]}. \quad (7)$$

The coarse-grained transition probabilities $\widetilde{W}_\lambda,\widetilde{W}_{\lambda^\dagger}$ are in general non-Markovian [we omitted the $P_0,P_{N\Delta t}$ dependence in the left hand side of Eq. (7)], therefore $\widetilde{\sigma}(\boldsymbol{x}_N)$ does not generally correspond to the actual heat entropy production in the reduced system. Nevertheless we adopt the formal definition (7), since it is the most natural extension of Eq. (4) to the non-Markovian process, and is an experimentally accessible quantity as far as x is an observable. Note that in the limit where the transition rates (5) and (6) become Markovian, which is the case of our main interest, $\widetilde{\sigma}(\boldsymbol{x}_N)$ represents the heat entropy production of the reduced system in the exact same sense that $\sigma(\boldsymbol{x}_N,\boldsymbol{y}_N)$ does for the whole system.

Let the irreversible entropy production be $\Sigma(\boldsymbol{x}_N,\boldsymbol{y}_N) := s(\boldsymbol{x}_N,\boldsymbol{y}_N) + \sigma(\boldsymbol{x}_N,\boldsymbol{y}_N)$, the coarse-grained irreversible entropy production be $\widetilde{\Sigma}(\boldsymbol{x}_N) := \widetilde{s}(\boldsymbol{x}_N) + \widetilde{\sigma}(\boldsymbol{x}_N)$, and the difference between them be

$$\Xi(\boldsymbol{x}_N,\boldsymbol{y}_N) := \Sigma(\boldsymbol{x}_N,\boldsymbol{y}_N) - \widetilde{\Sigma}(\boldsymbol{x}_N). \quad (8)$$

This $\Xi(\boldsymbol{x}_N,\boldsymbol{y}_N)$ is the entropy production that could not be caught when only the x dynamics is observed, thus we call it the hidden entropy production. Now for Σ and $\widetilde{\Sigma}$ the following holds:

$$\langle e^{-\Sigma(\boldsymbol{x}_N,\boldsymbol{y}_N)}\rangle_{\lambda,N} = 1, \quad (9)$$

$$\langle e^{-\widetilde{\Sigma}(\boldsymbol{x}_N)}\rangle_{\lambda,N} = 1. \quad (10)$$

Equation (9) is the well-known integral fluctuation theorem [16,17]. From this equality and Jensen's inequality we may show $\langle\Sigma(\boldsymbol{x}_N,\boldsymbol{y}_N)\rangle_{\lambda,N} = \langle s(\boldsymbol{x}_N,\boldsymbol{y}_N)\rangle_{\lambda,N} + \langle\sigma(\boldsymbol{x}_N,\boldsymbol{y}_N)\rangle_{\lambda,N} \geqslant 0$, the second law corresponding to the whole Markovian dynamics. Equation (10) is the integral fluctuation theorem for the coarse-grained system. The special case of Eq. (10) was mentioned in [18].

III. MAIN THEOREM

The first main result we present is that for the case where the variables x,y are time-reversal invariant $(x=\bar{x}, y=\bar{y})$, the following equality holds:

$$\langle e^{-\Xi(\boldsymbol{x}_N,\boldsymbol{y}_N)}\rangle_{\lambda,N} = 1. \quad (11)$$

This will be proved later. From Eq. (11) and Jensen's inequality, we see that the hidden entropy production is positive on average,

$$\langle\Xi(\boldsymbol{x}_N,\boldsymbol{y}_N)\rangle_{\lambda,N} \geqslant 0, \quad (12)$$

which means that the entropy production decreases due to the elimination of y.

Equality (11) is satisfied for the general two-variable (x,y) overdamped Langevin model, since in this case $x=\bar{x}$, $y=\bar{y}$. We have recently found [19] that nonzero hidden entropy production may exist even when a large time scale separation between the x and y motion justifies a closed Langevin description of x. In this example, we may directly prove the inequality (12) [19].

Here, let us present a more physical example, the Brownian heat engine model. We consider a one-dimensional underdamped Langevin dynamics [9],

$$\begin{aligned} \dot{x} &= p/m, \\ \dot{p} &= -\gamma p/m + F_{\lambda(t)}(x) + \sqrt{2\gamma T_{\lambda(t)}(x)}\xi(t), \end{aligned} \tag{13}$$

with x the Brownian particle's position (periodic boundary condition), p the momentum variable, and m the mass of the Brownian particle. $F(x) = f - \frac{\partial}{\partial x}U(x)$ is the general x dependent force with a constant load f and a spatially periodic potential $U(x)$. Setting the temperature $T(x)$ as periodic and inhomogeneous, one could extract work from the heat bath through the particle, hence this model serves as an autonomous heat engine [11,12]. In the usual experimental setup using a colloidal particle, the parameters in Eq. (13) justify the elimination of the fast variable p, therefore the effective dynamics obeys the overdamped equation,

$$\gamma\dot{x} = F_{\lambda(t)}(x) - \frac{\partial T_{\lambda(t)}(x)}{\partial x} + \sqrt{2\gamma T_{\lambda(t)}(x)} \circ \xi(t), \tag{14}$$

where $\circ$ denotes the Stratonovich product. Since $\bar{p} = -p$, Eq. (11) is not satisfied for the general parameter setup of (13). However, in the overdamped limit where the description Eq. (14) is justified, we may prove

$$\langle e^{-\Xi(x_\tau,p_\tau)}\rangle_{\lambda,\tau} = 1. \tag{15}$$

This specific equality was noted in [10]. Given a constant parameter [$\lambda(t) = \text{const}$], the hidden entropy production will have a steady rate,

$$\frac{\langle \Xi(\boldsymbol{x}_\tau,\boldsymbol{p}_\tau)\rangle_{\lambda,\tau}}{\tau} \xrightarrow{\tau\to\infty} \frac{1}{2\gamma}\left\langle \frac{1}{T(x)}\left(\frac{dT(x)}{dx}\right)^2\right\rangle_{\rm st}. \tag{16}$$

Such nonzero $\langle\Xi\rangle$ is the reason for the unattainability of Carnot efficiency in the Brownian heat engine [9].

IV. DETERMINISTIC DIFFUSION MODEL

One might expect from inequality (12) and the above Brownian heat engine example, that the decrease of entropy production is a general consequence of the Markovian limit coarse-graining. In this section we clarify that this is not the case by considering a deterministic Hamiltonian-like model, in which a probabilistic dynamics could be derived in the appropriate coarse-graining limit.

Our Markov chain model is an extension of the multibaker map [20]. The model is composed of many baker transformations that act on the nearest neighbor squares (Fig. 1). The variables $\xi,\eta \in [0,1]$ are the coordinates inside each unit area squares, and $r \in 1,2,\ldots,L$ is the label of those squares. We set a periodic boundary condition for r, and regard $r = L+1$ as $r = 1$ and $r = 0$ as $r = L$. We further introduce the "discretized velocity" variable, $v = +$ or $-$. The $v = +$ and $-$ systems are each composed of L squares, and are considered to be separated and noninteracting. The transition probability (deterministic map) from (r,ξ,η,v) to (r',ξ',η',v') in a unit time step is written as

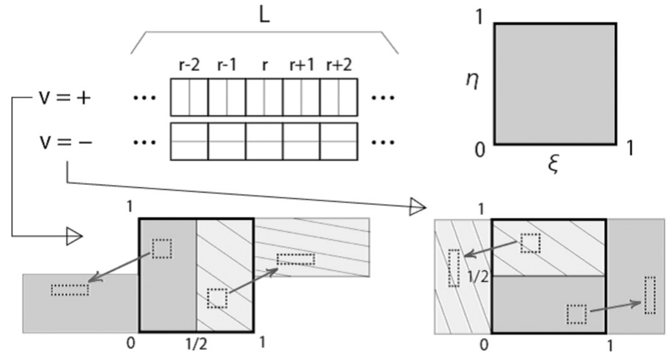

FIG. 1. Scheme of the multibaker map. The map considers the label of squares (r), two-dimensional coordinates inside the square (ξ,η), and the discretized velocity (v). The map is area-preserving and time-reversal invariant since the transition rules in the $v = +$ and $v = -$ systems are exactly opposite to each other [Eqs. (17), (19)].

$$W(r',\xi',\eta',v'|r,\xi,\eta,v) := \begin{cases} W^+(r',\xi',\eta'|r,\xi,\eta) & (v = v' = +) \\ W^-(r',\xi',\eta'|r,\xi,\eta) & (v = v' = -) \\ 0 & (\text{otherwise}) \end{cases} \tag{17}$$

with (see Fig. 1)

$$\begin{aligned} &W^+(r',\xi',\eta'|r,\xi,\eta) \\ &:= \theta\left(\tfrac{1}{2} - \xi\right)\delta_{r',r-1}\delta(\xi' - 2\xi)\delta\left(\eta' - \tfrac{\eta}{2}\right) \\ &\quad + \theta\left(\xi - \tfrac{1}{2}\right)\delta_{r',r+1}\delta(\xi' - 2\xi + 1)\delta\left(\eta' - \tfrac{\eta}{2} - \tfrac{1}{2}\right), \end{aligned} \tag{18}$$

$$\begin{aligned} &W^-(r',\xi',\eta'|r,\xi,\eta) \\ &:= \theta\left(\tfrac{1}{2} - \eta\right)\delta_{r',r+1}\delta\left(\xi' - \tfrac{\xi}{2}\right)\delta(\eta' - 2\eta) \\ &\quad + \theta\left(\eta - \tfrac{1}{2}\right)\delta_{r',r-1}\delta\left(\xi' - \tfrac{\xi}{2} - \tfrac{1}{2}\right)\delta(\eta' - 2\eta + 1), \end{aligned} \tag{19}$$

where $\theta(\cdot)$ is the Heaviside step function and $\delta(\cdot)$ is the Dirac delta function. Taking the time reversal of the variables as $(\bar{r},\bar{\xi},\bar{\eta},\bar{v}) = (r,\xi,\eta,-v)$, the model dynamics is completely time-reversal symmetric, $W(r',\xi',\eta',v'|r,\xi,\eta,v) = W(r,\xi,\eta,-v|r',\xi',\eta',-v')$.

Given an initial point (r_0,ξ_0,η_0,v), we gain a deterministic trajectory after N time steps $(\boldsymbol{r}_N,\boldsymbol{\xi}_N,\boldsymbol{\eta}_N,v)$. Writing the probability density function of (r,ξ,η,v) at time t as $P_t^v(r,\xi,\eta)$, we have

$$s(\boldsymbol{r}_N,\boldsymbol{\xi}_N,\boldsymbol{\eta}_N,v) = \log\frac{P_0^v(r_0,\xi_0,\eta_0)}{P_{N\Delta t}^v(r_N,\xi_N,\eta_N)} = 0, \tag{20}$$

because the phase space volume is conserved along the trajectories. We also find

$$\begin{aligned} \sigma(\boldsymbol{r}_N,\boldsymbol{\xi}_N,\boldsymbol{\eta}_N,v) &= \log\frac{W^v(\boldsymbol{r}_N,\boldsymbol{\xi}_N,\boldsymbol{\eta}_N|r_0,\xi_0,\eta_0)}{W^{-v}(\boldsymbol{r}_N^\dagger,\boldsymbol{\xi}_N^\dagger,\boldsymbol{\eta}_N^\dagger|r_N,\xi_N,\eta_N)} \\ &= 0, \end{aligned} \tag{21}$$

since we have set the dynamics in $v = -$ to be the complete time reversal of $v = +$. Therefore we have confirmed $\Sigma(\boldsymbol{r}_N,\boldsymbol{\xi}_N,\boldsymbol{\eta}_N,v) = s(\boldsymbol{r}_N,\boldsymbol{\xi}_N,\boldsymbol{\eta}_N,v) + \sigma(\boldsymbol{r}_N,\boldsymbol{\xi}_N,\boldsymbol{\eta}_N,v) = 0$. Note that this holds for any given initial density function $P_0^v(r_0,\xi_0,\eta_0)$.

Now we consider reducing the variables and only observing the discrete variable r. Then, the reduced irreversible entropy production $\widetilde{\Sigma}(\boldsymbol{r}_N)$ could be written using the "mesoscopic" probability distribution $\widetilde{P}_t(r) := \sum_v \int_0^1 d\xi \int_0^1 d\eta\, P_t^v(r,\xi,\eta)$, and transition probabilities $\widetilde{W}_t[r'|r,P_0^{\pm}(\cdot)] := \sum_v \int d\xi\, d\eta\, d\xi'\, d\eta'\, P_t^v(r,\xi,\eta) W^v(r',\xi',\eta'|r,\xi,\eta)/\widetilde{P}_t(r)$, $\widetilde{W}_t^{\dagger}[r'|r,P_{N\Delta t}^{\pm}(\cdot)] := \sum_v \int d\xi\, d\eta\, d\xi'\, d\eta'\, P_t^{v\dagger}(r,\xi,\eta) \times W^v(r',\xi',\eta'|r,\xi,\eta)/\widetilde{P}_t^{\dagger}(r)$. We find from Eq. (10) that for general initial distributions, the following holds:

$$\langle \Xi(\boldsymbol{r}_N,\boldsymbol{\xi}_N,\boldsymbol{\eta}_N,v)\rangle_N = -\langle \widetilde{\Sigma}(\boldsymbol{r}_N)\rangle_N \leqslant 0. \tag{22}$$

To confirm that the $\langle \widetilde{\Sigma}(\boldsymbol{r}_N)\rangle_N \neq 0$ case exists, we assume that $P_0^+(r,\xi,\eta)$ and $P_0^-(r,\xi,\eta)$ are smooth in the ξ and η directions, respectively, and define the time scale of the dynamics inside the squares as

$$\tau_{\xi,\eta} := \log \sup_{r,\xi,\eta} \left\{ \left| \frac{\partial}{\partial \xi} P_0^+(r,\xi,\eta)\right|, \left| \frac{\partial}{\partial \eta} P_0^-(r,\xi,\eta)\right| \right\}. \tag{23}$$

Then we may take, if L is sufficiently large, the "mesoscopic time scale" t^*, satisfying $\tau_{\xi,\eta} \ll t^* \ll \tau_r$. Here, τ_r ($\sim L^2$) is the typical time it takes for $\widetilde{P}_t(r)$ to become uniform ($\sim 1/L$). After this time t^*, the dynamics reduces to a simple random walk in the r direction,

$$\widetilde{W}_t, \widetilde{W}_t^{\dagger} \xrightarrow{t \geqslant t^*} \widetilde{W}(r'|r) = \tfrac{1}{2}\delta_{r',r+1} + \tfrac{1}{2}\delta_{r',r-1}. \tag{24}$$

Retaking the initial time $t = 0$ at this t^*, the average irreversible entropy production of the reduced dynamics satisfies

$$\langle \widetilde{\Sigma}(\boldsymbol{r}_N)\rangle_N = \sum_{r_0,r_1,\ldots,r_N} \widetilde{P}_0(r_0)\widetilde{W}(\boldsymbol{r}_N|r_0) \times \log \frac{\widetilde{P}_0(r_0)\widetilde{W}(\boldsymbol{r}_N|r_0)}{\widetilde{P}_{N\Delta t}(r_N)\widetilde{W}(\boldsymbol{r}_N^{\dagger}|r_N)} \geqslant 0. \tag{25}$$

Equality in (25) [and (22)] is achieved only when $\widetilde{P}_0(r_0)\widetilde{W}(\boldsymbol{r}_N|r_0) = \widetilde{P}_{N\Delta t}(r_N)\widetilde{W}(\boldsymbol{r}_N^{\dagger}|r_N)$ for all $\boldsymbol{r}_N$, that is, only when the given initial distribution is the equilibrium state, $\widetilde{P}_0(r) = 1/L$. Hence, we observe that the entropy production increases after the reduction in this model, as opposed to the case where Eq. (12) holds.

The inequality (22) states that the integral fluctuation theorem [Eq. (11)] does not hold in this model, except for the trivial case $\widetilde{\Sigma} = 0$ (equilibrium state). From Eq. (26) in the following section, we notice that the violation of Eq. (11) is due to the broken symmetry in the density function, $P_{N\Delta t}^+(\cdot) \neq P_{N\Delta t}^-(\cdot)$, which is valid for any $N > 0$ including $N \to \infty$ in this model. This is in clear contrast with the underdamped Langevin model, where the symmetry emerges in the overdamped (Markovian) limit [see Eq. (28)].

V. PROOFS

First we see that

$$\begin{aligned}
\langle e^{-\Xi(\boldsymbol{x}_N,\boldsymbol{y}_N)}\rangle_{\lambda,N}
&= \int d\boldsymbol{x}_N\, d\boldsymbol{y}_N\, P_\lambda(\boldsymbol{x}_N,\boldsymbol{y}_N) \\
&\quad\times \frac{P_{N\Delta t}(x_N,y_N) W_{\lambda^\dagger}(\boldsymbol{x}_N^{\dagger},\boldsymbol{y}_N^{\dagger}|\bar{x}_N,\bar{y}_N)\widetilde{P}_0(x_0)\widetilde{W}_\lambda[\boldsymbol{x}_N|x_0]}{P_0(x_0,y_0)W_\lambda(\boldsymbol{x}_N,\boldsymbol{y}_N|x_0,y_0)\widetilde{P}_{N\Delta t}(x_N)\widetilde{W}_{\lambda^\dagger}[\boldsymbol{x}_N^{\dagger}|\bar{x}_N]} \\
&= \int d\boldsymbol{x}_N\, d\boldsymbol{y}_N\, P_{\lambda^\dagger}(\boldsymbol{x}_N^{\dagger},\boldsymbol{y}_N^{\dagger})\frac{P_{N\Delta t}(x_N,y_N)}{P_{N\Delta t}(\bar{x}_N,\bar{y}_N)} \\
&\quad\times \frac{\widetilde{P}_0(x_0)\widetilde{W}_\lambda[\boldsymbol{x}_N|x_0]}{\widetilde{P}_{N\Delta t}(x_N)\widetilde{W}_{\lambda^\dagger}[\boldsymbol{x}_N^{\dagger}|\bar{x}_N]}.
\end{aligned} \tag{26}$$

For simplicity we omitted $P_0, P_{N\Delta t}$ in $\widetilde{W}_\lambda, \widetilde{W}_{\lambda^\dagger}$, respectively. Now if $\bar{x} = x$, $\bar{y} = y$, the right hand side of Eq. (26) reduces to

$$\begin{aligned}
&\int d\boldsymbol{x}_N\, d\boldsymbol{y}_N\, P_{\lambda^\dagger}(\boldsymbol{x}_N^{\dagger},\boldsymbol{y}_N^{\dagger})\frac{\widetilde{P}_0(x_0)\widetilde{W}_\lambda[\boldsymbol{x}_N|x_0]}{\widetilde{P}_{N\Delta t}(x_N)\widetilde{W}_{\lambda^\dagger}[\boldsymbol{x}_N^{\dagger}|x_N]} \\
&\quad= \int d\boldsymbol{x}_N\, \widetilde{P}_0(x_0)\widetilde{W}_\lambda[\boldsymbol{x}_N|x_0] = 1,
\end{aligned} \tag{27}$$

which is Eq. (11). In the case of the underdamped model (13), for a small parameter $\epsilon := \tau_p/\tau_x$ (τ_x is the fastest time scale of the motion in the x direction, and $\tau_p := m/\gamma$), we may show that the ratio $P_\tau(x,p)/P_\tau(x,-p)$ becomes close to unity, $1 + O(\epsilon)$, assuming that τ is large enough compared to τ_p. Since $\epsilon \to 0$ corresponds to the overdamped limit, Eq. (26) is now

$$\begin{aligned}
\langle e^{-\Xi(\boldsymbol{x}_\tau,\boldsymbol{p}_\tau)}\rangle_{\lambda,\tau} &= \int d\boldsymbol{x}_\tau\, d\boldsymbol{p}_\tau\, P_{\lambda^\dagger}(\boldsymbol{x}_\tau^{\dagger},\boldsymbol{p}_\tau^{\dagger}) \\
&\quad\times \frac{\widetilde{P}_0(x_0)\widetilde{W}_\lambda[\boldsymbol{x}_\tau|x_0]}{\widetilde{P}_\tau(x_\tau)\widetilde{W}_{\lambda^\dagger}[\boldsymbol{x}_\tau^{\dagger}|x_\tau]}[1 + O(\epsilon)] \xrightarrow{\epsilon\to 0} 1,
\end{aligned} \tag{28}$$

which is Eq. (15).

VI. REMARKS AND CONCLUSION

Recently, entropy production from accessible degrees of freedom was experimentally measured and analyzed in the interacting Brownian particle system [21]. Although their definition of "apparent entropy production" is different from our coarse-grained entropy production, the heat [Eq. (7)] and the hidden entropy productions [Eq. (8)] are also measurable quantities in their setup. Therefore, we claim that our main result Eq. (11) can be experimentally tested.

Next, we note the relation between our study and recent results on steady state thermodynamics [22]. In the overdamped limit of model (13), for the special case where Eq. (14) presents equilibrium dynamics [for instance, $F(x) = 0$], we observe that Eq. (11) is equivalent to the integral fluctuation theorem presented in [23] and [24]. This implies that in some cases the hidden entropy production is hidden inside the steady state dissipation in nonequilibrium dynamics. See [19] for related discussions.

Finally, we mention the interpretation of Eq. (11) in the context of information theory. Inequality (12) corresponds to the known inequality stating that the Kullback-Leibler (KL) divergence monotonically decreases by the Markov map of variables [25]. This correspondence follows from the fact that the ensemble averaged irreversible entropy production is equivalent to the KL divergence between the forward and backward path probabilities [18], and that the elimination of variables we discussed is a special Markov map $(x,y) \to x$. We may derive, for general Markov maps, an information theoretical counterpart of Eq. (11) [26]. However, the irreversible entropy production in the reduced dynamics does *not*, in general, correspond to the KL divergence after the Markov map, when time-reversal antisymmetric variables are included in the dynamics. This is why the hidden entropy production does not always obey the inequality (12), and therefore the irreversible entropy production may increase by coarse-graining.

To sum up, for a generic class of systems described only by time-reversal invariant variables, and for special cases where the final density is symmetric for all time-reversal antisymmetric variables, we found that the fluctuation theorem for the hidden entropy production [Eq. (11)] is satisfied and the entropy production always decreases. In the deterministic diffusion model, where the entropy production increases, Eq. (11) is violated due to the asymmetric density function. Since Eq. (11) has to be violated for the irreversible entropy production to increase after reducing variables, we find that such asymmetry of the distribution function must play an important role in the general derivation of stochastic processes from Hamiltonian dynamics. It is left for future studies to clarify this point in more physical kinetic equations.

ACKNOWLEDGMENTS

We thank T. Sagawa, M. Sano, S.-i. Sasa, and K. A. Takeuchi for fruitful discussions and reading of the manuscript. This work was supported by JSPS Research Fellowship.

[1] U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012).
[2] S. Toyabe, T. Okamoto, T. Watanabe-Nakayama, H. Taketani, S. Kudo, and E. Muneyuki, Phys. Rev. Lett. **104**, 198103 (2010).
[3] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano, Nat. Phys. **6**, 988 (2010).
[4] V. Blickle and C. Bechinger, Nat. Phys. **8**, 143 (2012).
[5] C. Bustamante, J. Liphardt, and F. Ritort, Phys. Today **58**(7), 43 (2005).
[6] K. Sekimoto, *Stochastic Energetics*, Lecture Notes in Physics (Springer, Berlin, 2010).
[7] N. G. van Kampen, Phys. Rep. **124**, 69 (1985).
[8] H. Spohn, Rev. Mod. Phys. **52**, 569 (1980).
[9] T. Hondou and K. Sekimoto, Phys. Rev. E **62**, 6021 (2000).
[10] A. Celani, S. Bo, R. Eichhorn, and E. Aurell, Phys. Rev. Lett. **109**, 260603 (2012).
[11] M. Büttiker, Z. Phys. B: Condens. Matter **68**, 161 (1987).
[12] R. Landauer, J. Stat. Phys. **53**, 233 (1988).
[13] J. Kurchan, J. Phys. A **31**, 3719 (1998).
[14] J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).
[15] C. Jarzynski, J. Stat. Phys. **98**, 77 (2000).
[16] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[17] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).
[18] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. Lett. **98**, 080602 (2007).
[19] Y. Nakayama and K. Kawaguchi, arXiv:1211.3805.
[20] P. Gaspard, J. Stat. Phys. **68**, 673 (1992).
[21] J. Mehl, B. Lander, C. Bechinger, V. Blickle, and U. Seifert, Phys. Rev. Lett. **108**, 220601 (2012).
[22] T. Hatano and S.-I. Sasa, Phys. Rev. Lett. **86**, 3463 (2001).
[23] T. Speck and U. Seifert, J. Phys. A **38**, L581 (2005).
[24] R. E. Spinney and I. J. Ford, Phys. Rev. Lett. **108**, 170603 (2012).
[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley Interscience, New York, 1991).
[26] M. Esposito, Phys. Rev. E **85**, 041125 (2012).