PHYSICAL REVIEW E **100**, 052127 (2019)

Three heats in a strongly coupled system and bath

Chulan Kwon,[1,*] Jaegon Um,[2,†] Joonhyun Yeo,[3] and Hyunggyu Park[4]
[1]*Department of Physics, Myongji University, Yongin, Gyeonggi-Do 17058, Korea*
[2]*BK21PLUS Physics Division, Pohang University of Science and Technology, Pohang 37673, Korea*
[3]*Department of Physics, Konkuk University, Seoul 05029, Korea*
[4]*School of Physics, Korea Institute for Advanced Study, Seoul 02455, Korea*

(Received 1 November 2018; revised manuscript received 10 September 2019; published 20 November 2019)

We investigate three kinds of heat produced in a system and a bath strongly coupled via an interaction Hamiltonian. By studying the energy flows between the system, the bath, and their interaction, we provide rigorous definitions of two types of heat, $Q_{\rm S}$ and $Q_{\rm B}$, from the energy loss of the system and the energy gain of the bath, respectively. This is in contrast to the equivalence of $Q_{\rm S}$ and $Q_{\rm B}$, which is commonly assumed to hold in the weak-coupling regime. The bath we consider is equipped with a thermostat which enables it to reach an equilibrium. We identify another kind of heat $Q_{\rm SB}$ from the energy dissipation of the bath into the superbath that provides the thermostat. We derive the fluctuation theorems (FTs) for the system variables and various heats, which are discussed in comparison with the FT for the total entropy production. We take an example of a sliding harmonic potential of a single Brownian particle in a fluid and calculate the three heats in a simplified model. These heats are found to equal, on average, in the steady state of energy, but show different fluctuations at all times.

DOI: 10.1103/PhysRevE.100.052127

I. INTRODUCTION

The nonequilibrium fluctuation theorem (FT) has been proven originally for deterministic systems [1–3], later for stochastic systems [4–10], and recently for quantum systems [11–14]. It takes into account thermodynamic quantities such as heat and work, which are continuously produced even in the steady state. Such quantities accumulated for a long time exhibit huge fluctuations around their means, which is especially prominent in small systems. Compared to work, heat is intriguing because it is interpreted as an energy exchange with a bath composed of infinitely many particles. By assuming the master equation or the Langevin equation, heat is found as a function of stochastic trajectories [15,16].

Recent studies, mostly quantum mechanical, have focused on a system *strongly* coupled with a bath [17–33]. In spite of extensive efforts, however, it is pointed out in Ref. [14] that a consistent definition of heat for the strong-coupling regime is currently not known and most of the studies are restricted to a specific initial (product) state. A proper means to treat the interaction energy or Hamiltonian between the system and the bath is still missing. This limitation is also present in classical approaches. In this study, we develop a theoretical framework to rigorously deal with the system-bath interaction for strongly coupled classical systems, which is expected to extend to quantum systems.

In this paper, we investigate a general nonequilibrium process in a classical system strongly coupled with a bath where three forms of heat appear in both system and bath. We find that a difference among the three heats is manifested in nonunique FTs and distinct distribution functions of the fluctuations. Results are explicitly confirmed for a specific example of nonequilibrium driven by a time-dependent protocol.

The remainder of this paper is organized as follows. In Sec. II, a strong-coupling model is introduced with constituent Hamiltonians and the equations of motion are given. In Sec. III, we show that three heats appear naturally and the first laws of thermodynamics for system and bath are expressed in terms of three heats and work. In Sec. IV, the fluctuation theorems of the entropy production and their variants with different heats are derived along with the corresponding thermodynamic inequalities. In Sec. V, we consider a specific example of a colloidal particle moving in a fluid under a sliding harmonic potential. We show that its simplified version can be mapped to an effective two-particle system, which can be solved analytically. The thermodynamic inequalities derived in Sec. IV are checked explicitly and the distinct features of three heat fluctuations are demonstrated. Finally, we summarize our results in Sec. VI.

II. MODEL FOR A STRONGLY COUPLED SYSTEM

We consider a general Hamiltonian for a system of particles coupled with a bath. The system variables are given by a collection of momenta $\vec{p}$ and positions $\vec{x}$, generally for multiparticle systems. Similarly, the bath variables are given by $\vec{p}_{\rm B}$ and $\vec{x}_{\rm B}$ for many bath particles. The Hamiltonian of the total system H is composed of three parts,

$$H_{\rm S} = \frac{\vec{p}^2}{2\mu} + U(\vec{x}, \lambda(t)), \quad H_{\rm B} = \frac{\vec{p}_{\rm B}^2}{2m} + U_{\rm B}(\vec{x}_{\rm B}),$$
$$H_{\rm I} = V(\vec{x}, \vec{x}_{\rm B}), \quad \text{with} \quad H = H_{\rm S} + H_{\rm B} + H_{\rm I}, \tag{1}$$

---

[*] ckwon@mju.ac.kr

[†] slung@postech.ac.kr

where $H_{\rm S}$, $H_{\rm B}$, and $H_{\rm I}$ stand for the Hamiltonians for the system, the bath, and their interaction, respectively. Note that the time-dependent protocol $\lambda(t)$ is prescribed only in the system potential U and the interaction Hamiltonian V is a pairwise potential between the system and bath particles. We take the same mass μ for all system particles and m for all bath particles, just for notational convenience.

We assume that the bath is equipped with a Langevin thermostat provided by an ideal superbath. Then, the classical equations of motion read

$$\begin{aligned} &\dot{\vec{x}} = \partial H_{\rm S}/\partial \vec{p}, \quad \dot{\vec{p}} = -\partial(H_{\rm S}+H_{\rm I})/\partial\vec{x},\\ &\dot{\vec{x}}_{\rm B} = \partial H_{\rm B}/\partial\vec{p}_{\rm B},\\ &\dot{\vec{p}}_{\rm B} = -\partial(H_{\rm B}+H_{\rm I})/\partial\vec{x}_{\rm B} - \gamma\vec{p}_{\rm B}/m + \vec{\xi}(t), \end{aligned} \tag{2}$$

with the white noise $\vec{\xi}(t)$ satisfying $\langle \xi_i(t)\xi_j(t')\rangle = 2\gamma\beta^{-1}\delta_{ij}\delta(t-t')$ for the inverse temperature β and the friction coefficient γ. In this way, the system and bath can be strongly coupled via (not so weak) interaction potential V, while the bath and superbath are weakly coupled via the Langevin thermostat.

III. THREE HEATS

The energy change rate for each Hamiltonian can be easily obtained from Eqs. (1) and (2) as

$$\begin{aligned} \frac{dH_{\rm S}}{dt} &= \frac{\partial U(t)}{\partial t} - \frac{\partial V}{\partial\vec{x}}\cdot\frac{\vec{p}}{\mu},\\ \frac{dH_{\rm B}}{dt} &= -\frac{\partial V}{\partial\vec{x}_{\rm B}}\cdot\frac{\vec{p}_{\rm B}}{m} - \left(\frac{\gamma\vec{p}_{\rm B}}{m} - \vec{\xi}\right)\cdot\frac{\vec{p}_{\rm B}}{m},\\ \frac{dH_{\rm I}}{dt} &= \frac{\partial V}{\partial\vec{x}}\cdot\frac{\vec{p}}{\mu} + \frac{\partial V}{\partial\vec{x}_{\rm B}}\cdot\frac{\vec{p}_{\rm B}}{m}, \end{aligned} \tag{3}$$

where we used the chain rules such as $dH_{\rm S}/dt = \partial H_{\rm S}/\partial t + (\partial H_{\rm S}/\partial\vec{p})\cdot\dot{\vec{p}} + (\partial H_{\rm S}/\partial\vec{x})\cdot\dot{\vec{x}}$ and so on. From the above relations, we can identify the rates of work and various types of heat. First, the rate of work induced by the time-dependent protocol $\lambda(t)$ in the system potential U is given as

$$\dot{W} = \frac{\partial U}{\partial t} = \frac{\partial U}{\partial\lambda}\dot{\lambda}, \tag{4}$$

which is known as Jarzynski's work [4,5].

We define the amount of energy loss of the system via the interaction with the bath as a heat $Q_{\rm S}$ and the accompanying energy gain of the bath as another heat $Q_{\rm B}$. In general, $Q_{\rm S}$ and $Q_{\rm B}$ are not identical due to the possible change of the interaction energy during any process. Since the bath is thermostatted to the superbath through the Langevin equation, the energy is dissipated from the bath into the superbath surrounding it, which is the third type of heat $Q_{\rm SB}$ considered in a recent study [31]. Then, the rates of the three heats are identified from Eq. (3) as

$$\dot{Q}_{\rm S} = \frac{\partial V}{\partial\vec{x}}\cdot\frac{\vec{p}}{\mu}, \quad \dot{Q}_{\rm B} = -\frac{\partial V}{\partial\vec{x}_{\rm B}}\cdot\frac{\vec{p}_{\rm B}}{m}, \quad \dot{Q}_{\rm SB} = \left(\frac{\gamma\vec{p}_{\rm B}}{m} - \vec{\xi}\right)\cdot\frac{\vec{p}_{\rm B}}{m}. \tag{5}$$

Finally, the energy-balance relations are expressed in terms of work and heats from Eq. (3) as

$$\begin{aligned} &\frac{dH_{\rm S}}{dt} = \dot{W} - \dot{Q}_{\rm S}, \quad \frac{dH_{\rm B}}{dt} = \dot{Q}_{\rm B} - \dot{Q}_{\rm SB},\\ &\frac{dH_{\rm I}}{dt} = \dot{Q}_{\rm S} - \dot{Q}_{\rm B}, \quad \frac{dH}{dt} = \dot{W} - \dot{Q}_{\rm SB}. \end{aligned} \tag{6}$$

The final equation for the total Hamiltonian is obtained by summing up the three other energy-balance equations, which obviously represents the thermodynamic first law of the total system with the input work energy and the output heat to the superbath. We emphasize that the superbath is necessary to keep the total system from incessant heating up by external driving, as expected in many realistic situations.

All other energy-balance equations are also easily understood. Note that work is done only on the system through the time-dependent protocol, and thus appears in the energy-balance equation for the system only. In the usual weak-coupling regime where the interaction Hamiltonian $H_{\rm I}$ is very small, one may recover the equivalence between two heats $\dot{Q}_{\rm S}$ and $\dot{Q}_{\rm B}$ [third equation in Eq. (6)]. However, this weak-coupling limit may become singular in some special cases, which will be discussed in Sec. V D.

The three heats are different by definition in Eq. (5). In a transient period, the three heats take different values, depending on the initial condition. In the long-time limit, if a (energy) steady state exists, we get $\langle dH_{\rm S}/dt\rangle = \langle dH_{\rm B}/dt\rangle = \langle dH_{\rm I}/dt\rangle = 0$, and, consequently, the average steady-state rates of the three heats become identical to each other, i.e., $\langle\dot{Q}_{\rm S}\rangle = \langle\dot{Q}_{\rm B}\rangle = \langle\dot{Q}_{\rm SB}\rangle = \langle\dot{W}\rangle$. However, even in the steady state, fluctuations of the three heats (and also work) do not have to be identical in general.

Next, we will show how the three heats behave differently with a focus on associated FTs and their fluctuations.

IV. FLUCTUATION THEOREMS FOR ENTROPY PRODUCTION AND THEIR VARIANTS

We examine the FT for our model. The total system (system and bath) is a many-particle system, of which the part (bath) is in contact with the ideal superbath. The standard integral FT for the total entropy production ΔS accumulated during a finite time interval still holds as

$$\langle e^{-\Delta S}\rangle = 1 \quad \text{with} \quad \Delta S = -\Delta\ln\rho + \beta Q_{\rm SB}, \tag{7}$$

where ρ is the probability distribution function (PDF) of the total system and the Boltzmann constant is set to be unity. The entropy production ΔS is simply the sum of the Shannon entropy change of the total system and the Clausius entropy production to the superbath [9], which indicates a weak coupling between the total system and the superbath. However, this trivial FT is *not* very informative in that the total system PDF depend on many degrees of freedom of the bath particles $(\vec{x}_{\rm B}, \vec{p}_{\rm B})$, which could not be inferred only from the system dynamics of $(\vec{x}, \vec{p})$. Thus, it would be useful to express the FT in terms of the system variables only with other energetic quantities such as the various heats and work discussed in the last section.

Let $\mathbf{q} = (\vec{x}, \vec{p}, \vec{x}_{\rm B}, \vec{p}_{\rm B})$ be a state vector of the total system with $\mathbf{q}_{\rm S} = (\vec{x}, \vec{p})$ for the system and $\mathbf{q}_{\rm B} = (\vec{x}_{\rm B}, \vec{p}_{\rm B})$ for the bath, respectively. The marginal PDF for the system is defined as

$$\rho_{\mathrm{S}}(\mathbf{q}_{\mathrm{S}}) = \mathrm{Tr}_{\mathrm{B}}\,\rho(\mathbf{q}), \tag{8}$$

where Tr_{B} is the trace over the bath variable $\mathbf{q}_{\mathrm{B}}$. Then, the Bayes' rule leads to

$$\rho(\mathbf{q}) = \rho_{\mathrm{S}}(\mathbf{q}_{\mathrm{S}})\rho(\mathbf{q}_{\mathrm{B}}|\mathbf{q}_{\mathrm{S}}), \tag{9}$$

where $\rho(\mathbf{q}_{\mathrm{B}}|\mathbf{q}_{\mathrm{S}})$ is the conditional PDF of $\mathbf{q}_{\mathrm{B}}$ for a given $\mathbf{q}_{\mathrm{S}}$. The marginal system PDF $\rho_{\mathrm{S}}(\mathbf{q}_{\mathrm{S}})$ can be measured experimentally or in numerical simulations when the system is sufficiently simple, e.g., a single- or a few-particle system. However, it is impossible to measure $\rho(\mathbf{q}_{\mathrm{B}}|\mathbf{q}_{\mathrm{S}})$ in realistic situations with many bath particles, except for some limiting cases where $\rho(\mathbf{q}_{\mathrm{B}}|\mathbf{q}_{\mathrm{S}})$ can be approximately estimated or guessed.

Let us imagine a *reference state* represented by the PDF

$$\widetilde{\rho}(\mathbf{q}) = \rho_{\mathrm{S}}(\mathbf{q}_{\mathrm{S}})\widetilde{\rho}(\mathbf{q}_{\mathrm{B}}|\mathbf{q}_{\mathrm{S}}), \tag{10}$$

where $\widetilde{\rho}(\mathbf{q}_{\mathrm{B}}|\mathbf{q}_{\mathrm{S}})$ is a guessed (estimated) conditional PDF of the bath state $\mathbf{q}_{\mathrm{B}}$ for a given system state $\mathbf{q}_{\mathrm{S}}$. Of course, $\widetilde{\rho}(\mathbf{q})$ would be different from the true PDF $\rho(\mathbf{q})$, in general. For convenience, we consider two typical reference states equipped by

$$\widetilde{\rho}(\mathbf{q}_{\mathrm{B}}|\mathbf{q}_{\mathrm{S}}) = Z_{\mathrm{B}}^{-1}e^{\beta\widetilde{H}_{\mathrm{S}}}e^{-\beta(H_{\mathrm{B}}+H_{\mathrm{I}})}, \tag{11a}$$

$$\widetilde{\rho}(\mathbf{q}_{\mathrm{B}}|\mathbf{q}_{\mathrm{S}}) = Z_{\mathrm{B}}^{-1}e^{-\beta H_{\mathrm{B}}}, \tag{11b}$$

where the bath equilibrium partition function $Z_{\mathrm{B}} = \mathrm{Tr}_{\mathrm{B}}e^{-\beta H_{\mathrm{B}}}$. We have an extra factor in the case of Eq. 11(a), called the *additional* system Hamiltonian $\widetilde{H}_{\mathrm{S}}(\mathbf{q}_{\mathrm{S}})$, which originates from the conditional probability normalization of $\mathrm{Tr}_{\mathrm{B}}\widetilde{\rho}(\mathbf{q}_{\mathrm{B}}|\mathbf{q}_{\mathrm{S}}) = 1$. Thus, we find

$$e^{-\beta\widetilde{H}_{\mathrm{S}}} = Z_{\mathrm{B}}^{-1}\mathrm{Tr}_{\mathrm{B}}e^{-\beta(H_{\mathrm{B}}+H_{\mathrm{I}})}. \tag{12}$$

One can see easily that $\widetilde{H}_{\mathrm{S}}$ vanishes when $H_{\mathrm{I}} = 0$, where two reference states coincide with each other.

The case given by Eq. 11(a) is a special type recently considered by Seifert [30], which describes a local equilibrium of a bath state for a given system state. If the total system is in equilibrium, then Eq. 11(a) is exact and the marginal system PDF becomes $\rho_{\mathrm{S}}(\mathbf{q}_{\mathrm{S}}) \propto e^{-\beta H_{\mathrm{S}}^{\mathrm{eff}}}$ with $H_{\mathrm{S}}^{\mathrm{eff}} = H_{\mathrm{S}} + \widetilde{H}_{\mathrm{S}}$, indicating that the strong coupling induces an additional term in the system Hamiltonian. The case given by Eq. 11(b) corresponds to the usual assumption of the product state of the system and bath in the weak-coupling case with the equilibrium bath.

The difference between the true and a reference state can be measured by the relative entropy between the two PDFs as $D(\rho||\widetilde{\rho}) = \ln[\rho/\widetilde{\rho}]$. Then, with Eq. (10), we get

$$\ln\rho = \ln\rho_{\mathrm{S}} + \ln\widetilde{\rho}(\mathbf{q}_{\mathrm{B}}|\mathbf{q}_{\mathrm{S}}) + D(\rho||\widetilde{\rho}). \tag{13}$$

Then, the total entropy production ΔS in Eq. (7) can be rewritten for each case as

$$\Delta S = -\Delta\ln\rho_{\mathrm{S}} + \beta(Q_{\mathrm{S}} - \Delta\widetilde{H}_{\mathrm{S}}) - \Delta D_{\mathrm{a}}, \tag{14a}$$

$$\Delta S = -\Delta\ln\rho_{\mathrm{S}} + \beta Q_{\mathrm{B}} - \Delta D_{\mathrm{b}}, \tag{14b}$$

where $\Delta D_{\mathrm{a,b}}$ are the relative entropy changes for the types 14(a) and 14(b), respectively. In this derivation, we used energy-balance relations in Eq. (6) such as $\Delta H_{\mathrm{B}} = Q_{\mathrm{B}} - Q_{\mathrm{SB}}$ and $\Delta(H_{\mathrm{B}} + H_{\mathrm{I}}) = Q_{\mathrm{S}} - Q_{\mathrm{SB}}$. Then, the thermodynamic second law, $\langle\Delta S\rangle \geqslant 0$, can be written with different average quantities as

$$\langle\Delta\ln\rho_{\mathrm{S}}\rangle + \beta(\langle Q_{\mathrm{S}}\rangle - \langle\Delta\widetilde{H}_{\mathrm{S}}\rangle) - \langle\Delta D_{\mathrm{a}}\rangle \geqslant 0, \tag{15a}$$

$$\langle\Delta\ln\rho_{\mathrm{S}}\rangle + \beta\langle Q_{\mathrm{B}}\rangle - \langle\Delta D_{\mathrm{b}}\rangle \geqslant 0, \tag{15b}$$

$$\langle\Delta\ln\rho\rangle + \beta\langle Q_{\mathrm{SB}}\rangle \geqslant 0, \tag{15c}$$

where the equality holds for nonthermostatted bath ($Q_{\mathrm{SB}} = 0$), due to $\langle\Delta\ln\rho\rangle = 0$ for the Liouville dynamics. Nevertheless, the above expressions still require the knowledge of the relative entropy change ΔD or the total PDF ρ, which cannot be accessible without knowing the true PDF $\rho(\mathbf{q}_{\mathrm{B}}|\mathbf{q}_{\mathrm{S}})$ of the bath.

One can get around this and extract some useful information when the initial state is not arbitrary but of our reference states given by Eq. 11(a) or Eq. 11(b). Moreover, these initial states can be easily prepared in real experiments. A similar strategy has been exploited in deriving the Jarzynski and Crooks FT for work with an equilibrium initial state for the total system [4–6].

We now derive alternative FTs depending on the choice of initial states. Consider a quantity ΔA as

$$\Delta A \equiv -\Delta\ln\rho_{\mathrm{S}} + \beta(Q_{\mathrm{S}} - \Delta\widetilde{H}_{\mathrm{S}}), \tag{16}$$

which appears in Eq. (14) as a part of ΔS. Note that this quantity does not require the knowledge of the bath PDF. With the initial condition prepared with the reference state 11(a) and for a finite time interval $t = [0, \tau]$, we find an alternative FT as

$$\begin{aligned}\langle e^{-\Delta A}\rangle_{\mathrm{a}} &= \int D\mathbf{q}(t)\,e^{-\Delta A}\,\Pi[\mathbf{q}(t);\lambda(t)]\frac{\rho_{\mathrm{S}}(0)e^{-\beta[H_{\mathrm{B}}(0)+H_{\mathrm{I}}(0)]}}{Z_{\mathrm{B}}e^{-\beta\widetilde{H}_{\mathrm{S}}(0)}}\\ &= \int D\mathbf{q}_{\mathrm{R}}(t)\,\Pi[\mathbf{q}_{\mathrm{R}}(t);\lambda_{\mathrm{R}}(t)]\frac{\rho_{\mathrm{S}}(\tau)e^{-\beta[H_{\mathrm{B}}(\tau)+H_{\mathrm{I}}(\tau)]}}{Z_{\mathrm{B}}e^{-\beta\widetilde{H}_{\mathrm{S}}(\tau)}}\\ &= 1,\end{aligned} \tag{17}$$

where $\langle\cdot\rangle_{\mathrm{a}}$ is the path average with the initial state of reference type 11(a) and $\Pi[\mathbf{q}(t);\lambda(t)]$ is the standard conditional probability for the path $\{\mathbf{q}(t), t = [0, \tau]\}$ under the protocol $\lambda(t)$ [10,34]. The second equality comes from the well-known Schnakenberg relation $\Pi[\mathbf{q}(t);\lambda(t)]/\Pi[\mathbf{q}_{\mathrm{R}}(t);\lambda_{\mathrm{R}}(t)] = e^{\beta Q_{\mathrm{SB}}}$ [7,8,15,34] with the conditional probability $\Pi[\mathbf{q}_{\mathrm{R}}(t);\lambda_{\mathrm{R}}(t)]$ for the time-reverse path $\{\mathbf{q}_{\mathrm{R}}(t), t = [\tau, 0]\}$ under the time-reverse protocol $[\lambda_{\mathrm{R}}(t) = \lambda(\tau - t)]$ and the unit Jacobian between two path-integral variables. In addition, we used the energy-balance relation $\Delta(H_{\mathrm{B}} + H_{\mathrm{I}}) = Q_{\mathrm{S}} - Q_{\mathrm{SB}}$ again from Eq. (6). The final equality comes from the probability normalization because the second integral represents the sum of all possible path probabilities in the time-reverse process with its initial state of the same reference type 11(a).

Similarly, we find another alternative FT with the initial condition prepared with the reference state 11(b) as

$$\Delta B \equiv -\Delta\ln\rho_{\mathrm{S}} + \beta Q_{\mathrm{B}}, \quad \langle e^{-\Delta B}\rangle_{\mathrm{b}} = 1. \tag{18}$$

The FT for ΔA in Eq. (17) has been recently found by Seifert [30] in the case without the superbath, and the FT for ΔB has been known awhile for quantum systems [32]. As the FT variables in both cases concern the system dynamics only with energy (heat) variables, these FTs are much more useful than the FT for the total entropy production in Eqs. (7) and (14). However, the choice of the initial condition is crucial, just like in the Jarzynski relation for work [4,5].

The corresponding inequalities are given as

$$R_{\mathrm{A}} = \langle -\Delta \ln \rho_{\mathrm{S}} + \beta(Q_{\mathrm{S}} - \Delta \widetilde{H}_{\mathrm{S}}) \rangle_{\mathrm{a}} \geqslant 0, \tag{19a}$$

$$R_{\mathrm{B}} = \langle -\Delta \ln \rho_{\mathrm{S}} + \beta Q_{\mathrm{B}} \rangle_{\mathrm{b}} \geqslant 0. \tag{19b}$$

Theses inequalities can also be directly derived from Eq. (15) by setting the initial state identical to the reference state, i.e., $\rho(0) = \widetilde{\rho}$. Then the relative entropy $D(\rho||\widetilde{\rho})$ is simply zero at the initial time ($t = 0$). The relative entropy at later time t may not be zero because the true PDF $\rho(t)$ does not maintain its form of reference states in general as soon as the evolution starts. Therefore, we find $R_{\mathrm{A}} - \langle \Delta S \rangle_{\mathrm{a}} = \langle D_{\mathrm{a}}(\tau) \rangle_{\mathrm{a}} \geqslant 0$, where the last inequality comes from the well-known non-negative property of a relative entropy. Similarly, $R_{\mathrm{B}} - \langle \Delta S \rangle_{\mathrm{b}} = \langle D_{\mathrm{b}}(\tau) \rangle_{\mathrm{b}} \geqslant 0$. Thus, the above inequalities in Eq. (17) are obtained from the inequalities for the total entropy production in Eq. (15). It is interesting to note that R_{A} and R_{B} are useful in practice at the cost of information loss, which yields a looser bound by the amount of the relative entropy at the final time than the exact bound given by the total entropy production.

One should also notice that R_{A} and R_{B} do not necessarily increase with the interval time τ, i.e., they may go up and down with increasing τ even though their non-negativity is maintained. This is also simply because the true PDF would deviate from the reference form after the evolution starts. The similar behavior is found for the dissipated work in the Jarzynski FT and also some quantum FT variables [31]. We will show these interesting properties explicitly for an exactly solvable simple model in the next section.

V. NONEQUILIBRIUM MOTION UNDER A SLIDING POTENTIAL

We take a concrete example to demonstrate the difference among the three heats. Consider a Brownian colloidal particle submerged in a fluid bath. This colloid interacts with bath particles nearby through a finite-range interaction. Perturbed bath particles via interaction with the colloid relax fast into equilibrium and new bath particles begin to interact as the colloid moves through the bath. For an analytic approach, we mimic this situation by considering only a finite number n of bath particles moving along with the colloid through strong harmonic interactions [35]. All other noninteracting bath particles are in equilibrium and will not be taken into account.

For simplicity, we only consider the one-dimensional model and assume no bath potential ($U_{\mathrm{B}} = 0$). The system and bath state are given by $\mathbf{q}_{\mathrm{S}} = (x, p)$ and $\mathbf{q}_{\mathrm{B}} = (x_1, \ldots, x_n, p_1, \ldots, p_n)$, respectively. Note that we dropped the state variables of all other bath particles which do not interact with the colloid. The interaction potential between the colloid and the ith bath particles is chosen as $V_i = \kappa(x - x_i)^2/2$, which is long ranged enough to keep interacting bath particles near the colloid.

In order to study not only equilibrium but also nonequilibrium steady states, we apply a time-dependent protocol on the colloid by introducing a sliding harmonic potential with a constant velocity u given by $U(x, \lambda(t)) = k(x - ut)^2/2$ with $\lambda(t) = ut$. This protocol for a Brownian particle has been extensively studied experimentally and theoretically [36–42] for a single-particle Langevin system.

The total Hamiltonian is given as

$$H = \frac{p^2}{2\mu} + \frac{k}{2}(x - ut)^2 + \sum_{i=1}^{n} \frac{p_i^2}{2m} + \sum_{i=1}^{n} \frac{\kappa}{2}(x - x_i)^2, \tag{20}$$

where H_{S}, H_{B}, and H_{I} can be easily identified. The equations of motion are obtained from Eq. (2) as

$$\begin{aligned} \dot{x} &= \frac{p}{\mu}, \ \dot{p} = -k(x - ut) - \kappa \sum_{i=1}^{n} (x - x_i), \\ \dot{x}_i &= \frac{p_i}{m}, \ \dot{p}_i = -\kappa(x_i - x) - \gamma \frac{p_i}{m} + \xi_i, \end{aligned} \tag{21}$$

where the interacting bath particle index i runs from 1 to n.

A. Decoupling by coordinate transformation

Due to the simple linear nature of the equations of motions in Eq. (21), we can decompose the $(n + 1)$-particle coupled dynamics into decoupled effective dynamics. To this end, we introduce the center-of-mass (CM) coordinates for n interacting bath particles as

$$x_{\mathrm{cm}} \equiv \frac{1}{n} \sum_{i=1}^{n} x_i, \quad p_{\mathrm{cm}} \equiv \sum_{i=1}^{n} p_i, \tag{22}$$

leading to $p_{\mathrm{cm}} = (nm)\dot{x}_{\mathrm{cm}}$. All other reduced (relative) bath coordinates are defined as

$$x_i' \equiv \frac{\sum_{j=1}^{i} x_j - i x_{i+1}}{\sqrt{i(i+1)}}, \quad p_i' \equiv \frac{\sum_{j=1}^{i} p_j - i p_{i+1}}{\sqrt{i(i+1)}}, \tag{23}$$

for $i = 1, \ldots, n - 1$.

Then, the equations of motion for the system and a representative bath particle at the CM coordinates are given as

$$\begin{aligned} \dot{x} &= \frac{p}{\mu}, \ \dot{p} = -k(x - ut) - n\kappa(x - x_{\mathrm{cm}}), \\ \dot{x}_{\mathrm{cm}} &= \frac{p_{\mathrm{cm}}}{nm}, \ \dot{p}_{\mathrm{cm}} = -n\,\kappa(x_{\mathrm{cm}} - x) - n\,\gamma \frac{p_{\mathrm{cm}}}{nm} + \xi_{\mathrm{cm}}, \end{aligned} \tag{24}$$

where $\xi_{\mathrm{cm}} = \sum_i \xi$ is the total noise acting on the CM, satisfying $\langle \xi_{\mathrm{cm}}(t) \xi_{\mathrm{cm}}(t') \rangle = 2n\gamma\beta^{-1}\delta(t - t')$. This implies that the colloid effectively interacts with a single bath particle of mass nm at the CM coordinate, subject to the friction coefficient $n\gamma$ and the harmonic interaction strength $n\kappa$. The equations of motion for the other $n - 1$ reduced bath coordinates are given as

$$\dot{x}_i' = \frac{p_i'}{m}, \quad \dot{p}_i' = -\kappa x_i' - \gamma \frac{p_i'}{m} + \xi_i', \tag{25}$$

where $\xi_i' \equiv [\sum_{j=1}^{i} \xi_j - i\xi_{i+1}]/\sqrt{i(i+1)}$, satisfying $\langle \xi_i'(t) \xi_j'(t') \rangle = 2\gamma\beta^{-1}\delta_{ij}\delta(t - t')$. One can also easily show that ξ_{CM} and ξ_i' are not correlated. As evidenced in Eq. (25), each reduced bath coordinate is independent from each other and also from the CM and colloid coordinate. Thus, the total system is completely decoupled into an *interacting* two-particle system (colloid and CM) and $(n-1)$ noninteracting single-particle system (reduced bath coordinate).

The Hamiltonian for the interacting two-particle system $H^{(2)}$ can be written as

$$H_{\mathrm{S}}^{(2)} = \frac{p^2}{2\mu} + \frac{k}{2}(x-ut)^2, \quad H_{\mathrm{B}}^{(2)} = \frac{p_{\mathrm{cm}}^2}{2nm},$$
$$H_{\mathrm{I}}^{(2)} = \frac{n\kappa}{2}(x-x_{\mathrm{cm}})^2, \quad \text{with } H^{(2)} = H_{\mathrm{S}}^{(2)} + H_{\mathrm{B}}^{(2)} + H_{\mathrm{I}}^{(2)}. \tag{26}$$

For the reduced bath system, we have the Hamiltonian $H^{(n-1)}$ for $n-1$ particles as

$$H_{\mathrm{B}}^{(n-1)} = \sum_{i=1}^{n-1} \frac{p_i'^2}{2m}, \quad H_{\mathrm{I}}^{(n-1)} = \sum_{i=1}^{n-1} \frac{\kappa}{2} x_i'^2, \tag{27}$$

with

$$H^{(n-1)} = H_{\mathrm{B}}^{(n-1)} + H_{\mathrm{I}}^{(n-1)},$$

where the term with the subscript I comes from the interaction term in the original Hamiltonian of Eq. (20). The reduced bath system is a collection of $n-1$ noninteracting harmonic oscillators. In this description, we have one colloidal particle for the system, one (CM) particle for the interacting bath, and $n-1$ (reduced) particles for the noninteracting bath. We note that the Hamiltonians in different representations have relations as

$$H_{\mathrm{S}} = H_{\mathrm{S}}^{(2)}, \qquad H_{\mathrm{B}} = H_{\mathrm{B}}^{(2)} + H_{\mathrm{B}}^{(n-1)},$$
$$H_{\mathrm{I}} = H_{\mathrm{I}}^{(2)} + H_{\mathrm{I}}^{(n-1)}, \quad \text{with } H = H^{(2)} + H^{(n-1)}. \tag{28}$$

B. Interacting two-particle system

All essential and nontrivial features are embedded in the interacting two-particle dynamics with the Hamiltonian $H^{(2)}$ in Eq. (26). We denote a state vector of this two-particle system as $\vec{q} = (x, p, x_{\mathrm{cm}}, p_{\mathrm{cm}})$. It is convenient to introduce a shifted state vector $\vec{r} = (x-ut, p, x_{\mathrm{cm}}-ut, p_{\mathrm{cm}})$. Then, Eq. (24) can be written as

$$\dot{\vec{r}} = -\mathsf{F}\cdot\vec{r} - \vec{u} + \vec{\eta}, \tag{29}$$

where F is a 4×4 matrix, and $\vec{u}$, $\vec{\eta}$ are four-component vectors, given as

$$\mathsf{F} = \begin{pmatrix} \mathsf{F}_{\mathrm{S}} & \mathsf{F}_{\mathrm{I}} \\ \mathsf{F}_{\mathrm{I}} & \mathsf{F}_{\mathrm{B}} \end{pmatrix}, \quad \vec{u} = \begin{pmatrix} u \\ 0 \\ u \\ 0 \end{pmatrix}, \quad \vec{\eta} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \xi_{\mathrm{cm}} \end{pmatrix},$$

with the 2×2 block matrices given as

$$\mathsf{F}_{\mathrm{S}} = \begin{pmatrix} 0 & -1/\mu \\ k+n\kappa & 0 \end{pmatrix}, \quad \mathsf{F}_{\mathrm{B}} = \begin{pmatrix} 0 & -1/(nm) \\ n\kappa & \gamma/m \end{pmatrix},$$
$$\mathsf{F}_{\mathrm{I}} = \begin{pmatrix} 0 & 0 \\ -n\kappa & 0 \end{pmatrix}.$$

Note that $\vec{r} = \vec{q} - \vec{u}t$ and Eq. (29) describe an equilibrium dynamics at $\vec{u} = 0$. Single-particle systems (overdamped and underdamped) driven by a sliding harmonic potential have been studied extensively in the literature [39–43]. We use a similar strategy to exactly solve our two-particle interacting system in this study.

First, we decompose the shifted state vector $\vec{r}$ into the deterministic part $\vec{d} = \langle\vec{r}\rangle$ and the stochastic part $\vec{z} = \vec{r} - \vec{d}$. Then, we get the two decoupled equations as

$$\dot{\vec{d}} = -\mathsf{F}\cdot\vec{d} - \vec{u}, \quad \dot{\vec{z}} = -\mathsf{F}\cdot\vec{z} + \vec{\eta}. \tag{30}$$

The deterministic part can be easily solved as

$$\vec{d}(t) = -\mathsf{F}^{-1}(\mathsf{I} - e^{-\mathsf{F}t})\cdot\vec{u}, \tag{31}$$

where we take the initial condition $\vec{d}(0) = \vec{0}$ for simplicity, implying $\langle x(0)\rangle = \langle p(0)\rangle = \langle x_{\mathrm{cm}}(0)\rangle = \langle p_{\mathrm{cm}}(0)\rangle = 0$.

The stochastic part describes the famous multivariate Ornstein-Uhlenbeck process, which has been well studied [44–46]. In our case, F is trivial in the sense that it does not contain any nonequilibrium source, so we expect a simple equilibrium relaxation process. The PDF for its evolution is given as

$$P(\vec{z}, t) = \sqrt{\frac{\beta^4 \det \mathsf{A}_t}{(2\pi)^4}} \exp\left[-\frac{\beta}{2}\vec{z}\cdot\mathsf{A}_t\cdot\vec{z}\right], \tag{32}$$

with

$$\mathsf{A}_t^{-1} = \mathsf{A}_{\mathrm{eq}}^{-1} - e^{-\mathsf{F}t}\left(\mathsf{A}_{\mathrm{eq}}^{-1} - \mathsf{A}_0^{-1}\right)e^{-\mathsf{F}^{\mathrm{t}}t}, \tag{33}$$

where A_{eq} and A_0 correspond to the matrices for the equilibrium distribution and an initial Gaussian distribution, respectively [44,46]. As expected, the PDF approaches the Boltzmann distribution as $t\to\infty$ with A_{eq}, which can be decomposed as

$$\mathsf{A}_{\mathrm{eq}} = \mathsf{A}_{\mathrm{S}} + \mathsf{A}_{\mathrm{B}} + \mathsf{A}_{\mathrm{I}}, \tag{34}$$

where

$$\mathsf{A}_{\mathrm{S}} = \begin{pmatrix} \mathsf{S} & \mathsf{0} \\ \mathsf{0} & \mathsf{0} \end{pmatrix} \text{ with } \mathsf{S} = \begin{pmatrix} k & 0 \\ 0 & 1/\mu \end{pmatrix},$$
$$\mathsf{A}_{\mathrm{B}} = \begin{pmatrix} \mathsf{0} & \mathsf{0} \\ \mathsf{0} & \mathsf{B} \end{pmatrix} \text{ with } \mathsf{B} = \begin{pmatrix} 0 & 0 \\ 0 & 1/(nm) \end{pmatrix}, \tag{35}$$
$$\mathsf{A}_{\mathrm{I}} = \begin{pmatrix} \mathsf{K} & -\mathsf{K} \\ -\mathsf{K} & \mathsf{K} \end{pmatrix} \text{ with } \mathsf{K} = \begin{pmatrix} n\kappa & 0 \\ 0 & 0 \end{pmatrix},$$

and $\mathsf{0}$ is the 2×2 null matrix. It would be very useful later to realize that each component corresponds to one of the Hamiltonian components in Eq. (26) such as

$$H^{(2)} = \tfrac{1}{2}\vec{r}\cdot\mathsf{A}_{\mathrm{eq}}\cdot\vec{r}, \quad H_{\mathrm{S}}^{(2)} = \tfrac{1}{2}\vec{r}\cdot\mathsf{A}_{\mathrm{S}}\cdot\vec{r},$$
$$H_{\mathrm{B}}^{(2)} = \tfrac{1}{2}\vec{r}\cdot\mathsf{A}_{\mathrm{B}}\cdot\vec{r}, \quad H_{\mathrm{I}}^{(2)} = \tfrac{1}{2}\vec{r}\cdot\mathsf{A}_{\mathrm{I}}\cdot\vec{r}. \tag{36}$$

The PDF for the two-particle system $\rho^{(2)}$ is given as

$$\rho^{(2)}(\vec{q}, t) = P(\vec{r} - \vec{d}(t), t) = P(\vec{q} - \vec{u}t - \vec{d}(t), t), \tag{37}$$

where the function P is given in Eq. (32). As t goes to ∞, the PDF reaches the Boltzmann distribution with its center moving as $\langle\vec{q}\rangle = \vec{u}t + \vec{d}(t)$. In the long-time limit, $\vec{d}(t)$ saturates [see Eq. (31)], so the PDF center moves with the constant velocity u.

The PDF of the total system including $n-1$ noninteracting bath particles is simply given by product as

$$\rho(\vec{q},t)=\rho^{(2)}(\vec{q},t)\prod_{i=1}^{n-1}\rho_{\mathrm{HO}}(x_i',p_i',t), \tag{38}$$

where ρ_{HO} is a distribution function for a single harmonic oscillator in Eq. (27).

The marginal system PDF $\rho_{\mathrm{S}}(\vec{q}_{\mathrm{S}},t)$ with $\vec{q}_{\mathrm{S}}=(x,p)$ can be easily obtained by tracing out all the bath coordinates in Eq. (38) as

$$\rho_{\mathrm{S}}(\vec{q}_S,t)=\frac{\sqrt{\beta^2\det\widehat{\mathsf{S}}_t}}{2\pi}\exp\left[-\frac{\beta}{2}\vec{z}_{\mathrm{S}}\cdot\widehat{\mathsf{S}}_t\cdot\vec{z}_{\mathrm{S}}\right], \tag{39}$$

where $\vec{z}_{\mathrm{S}}$ is the stochastic part of $\vec{q}_{\mathrm{S}}$ defined as $\vec{z}_{\mathrm{S}}=[x-ut-d_1(t),p-d_2(t)]$ with the subscript i denoting the ith vector component of $\vec{d}$ in Eq. (31). The covariance matrix $[\widehat{\mathsf{S}}_t^{-1}]_{ij}=\langle z_i(t)z_j(t)\rangle_{\rho_{\mathrm{S}}}$ for $i,j=1,2$ [44,46] should be identical to the covariant matrix $[\mathsf{A}_t^{-1}]_{ij}$ for the two-particle system in Eq. (32). Note that the PDF for all other noninteracting harmonic oscillators is traced out without affecting the marginal system PDF.

C. Average work and heats

The expression for the nonequilibrium work can be obtained from Eq. (4) as

$$W=\int_0^\tau dt\,\frac{\partial U}{\partial\lambda}\dot{\lambda}=-ku\int_0^\tau dt(x-ut). \tag{40}$$

Note that $x-ut$ is the first component of the shifted state vector $\vec{r}=\vec{z}+\vec{d}$. Then, the average work can be calculated easily as

$$\begin{aligned}\langle W\rangle&=-ku\int_0^\tau dt\,d_1(t)\\&=-ku[\{\mathsf{F}^{-1}\tau-(\mathsf{F}^{-1})^2(\mathsf{I}-e^{-\mathsf{F}\tau})\}\cdot\vec{u}]_1,\end{aligned} \tag{41}$$

where we used $d_1(t)$ from Eq. (31) and $\langle\vec{z}\rangle=0$ from Eq. (32). In the long-time limit ($\tau\to\infty$), the first term dominates and we get the constant average work production rate as

$$\langle\dot{W}\rangle\simeq n\gamma u^2. \tag{42}$$

We will show later that the heat production rate is also a constant and equal to the work production rate in the long-time limit. This implies that there exists the *energy-balance* steady state where the work production rate and heat production rate can be balanced with no internal energy change on average, even though there is no true steady state from the PDF point of view.

We utilize the energy-balance equation in Eq. (6) to get the expression for the three heats as

$$Q_\alpha=W-\Delta\widehat{H}_\alpha, \tag{43}$$

where $\widehat{H}_\alpha=H_{\mathrm{S}},H_{\mathrm{S}}+H_{\mathrm{I}},H$ for $\alpha=\mathrm{S},\mathrm{B},\mathrm{SB}$, respectively. We can rewrite $\widehat{H}_\alpha$ in decoupled coordinates using the Hamiltonian relations in Eq. (28) as

$$\widehat{H}_\alpha=\widehat{H}_\alpha^{(2)}+\widehat{H}_\alpha^{(n-1)}, \tag{44}$$

with

$$\begin{aligned}&\widehat{H}_{\mathrm{S}}^{(2)}=H_{\mathrm{S}}^{(2)},\quad \widehat{H}_{\mathrm{S}}^{(n-1)}=0,\\&\widehat{H}_{\mathrm{B}}^{(2)}=H_{\mathrm{S}}^{(2)}+H_{\mathrm{I}}^{(2)},\quad \widehat{H}_{B}^{(n-1)}=H_{\mathrm{I}}^{(n-1)},\\&\widehat{H}_{\mathrm{SB}}^{(2)}=H^{(2)},\quad \widehat{H}_{\mathrm{SB}}^{(n-1)}=H^{(n-1)}.\end{aligned} \tag{45}$$

Note that $\widehat{H}_\alpha^{(2)}$ and $\widehat{H}_\alpha^{(n-1)}$ are completely independent of each other with different coordinates.

In order to obtain the average values of the heats $\langle Q_\alpha\rangle$, we only need to calculate the average Hamiltonian change $\langle\Delta\widehat{H}_\alpha\rangle$ because the average work was already obtained in Eq. (41). First, we calculate the contribution from the two-particle Hamiltonian, i.e., $\langle\Delta\widehat{H}_\alpha^{(2)}\rangle$. Using Eq. (36), $\widehat{H}_\alpha^{(2)}$ can be rewritten in the form of $\frac{1}{2}\vec{r}\cdot\widehat{\mathsf{A}}_\alpha\cdot\vec{r}$, with

$$\widehat{\mathsf{A}}_{\mathrm{S}}=\mathsf{A}_{\mathrm{S}}\,,\quad \widehat{\mathsf{A}}_{\mathrm{B}}=\mathsf{A}_{\mathrm{S}}+\mathsf{A}_{\mathrm{I}}\,,\quad \widehat{\mathsf{A}}_{\mathrm{SB}}=\mathsf{A}_{\mathrm{eq}}. \tag{46}$$

Then, we get

$$\begin{aligned}\big\langle\Delta\widehat{H}_\alpha^{(2)}\big\rangle&=\Delta\big\langle\tfrac{1}{2}[\vec{z}(t)+\vec{d}(t)]\cdot\widehat{\mathsf{A}}_\alpha\cdot[\vec{z}(t)+\vec{d}(t)]\big\rangle\\&=\tfrac{1}{2}\vec{d}(\tau)\cdot\widehat{\mathsf{A}}_\alpha\cdot\vec{d}(\tau)-\tfrac{1}{2}\mathrm{Tr}\widehat{\mathsf{A}}_\alpha\big(\mathsf{A}_\tau^{-1}-\mathsf{A}_0^{-1}\big),\end{aligned} \tag{47}$$

where we used $\vec{d}(0)=0$ and the covariance matrix $[\mathsf{A}_t^{-1}]_{ij}=\langle z_i(t)z_j(t)\rangle$. Together with the expression for the average work in Eq. (41), we have the exact expression for $\langle Q_{\mathrm{S}}\rangle$ because $\widehat{H}_{\mathrm{S}}^{(n-1)}=0$. With the further assumption that the reduced bath is in equilibrium with $H^{(n-1)}$ at $t=0$, then $\langle\Delta\widehat{H}_\alpha^{(n-1)}\rangle=0$, and thus we have the exact expressions for the other two heats, $\langle Q_{\mathrm{B}}\rangle$ and $\langle Q_{\mathrm{SB}}\rangle$.

In the long-time limit, the internal energy changes in Eq. (47) saturate and do not increase linearly in time τ. Thus, the leading term in τ for all average heats $\langle Q_\alpha\rangle$ comes from the work production $\langle W\rangle$, leading to the expected result for the work and heat production rate in the long-time limit as

$$\langle\dot{W}\rangle\simeq\langle\dot{Q}_{\mathrm{S}}\rangle\simeq\langle\dot{Q}_{\mathrm{B}}\rangle\simeq\langle\dot{Q}_{\mathrm{SB}}\rangle\simeq n\gamma u^2. \tag{48}$$

As we have the exact expressions for the total PDF in Eq. (38) and $\langle Q_{\mathrm{SB}}\rangle=\langle W\rangle-\langle\Delta\widehat{H}_{\mathrm{SB}}\rangle$ with Eqs. (41) and (47), we can exactly calculate the time evolution of the average total entropy production $\langle\Delta S\rangle$ with a given initial condition. Furthermore, we can also calculate the time evolution of $R_{\mathrm{A}}=\langle\Delta A\rangle_a$ in Eq. (17), which is expected to behave nonmonotonically with time, as discussed in Sec. IV. To see this behavior explicitly, we take a simple initial PDF as

$$\rho(\mathbf{q},0)\propto\exp[-\beta' H_{\mathrm{S}}-\beta(H_{\mathrm{B}}+H_{\mathrm{I}})]. \tag{49}$$

As $H_{\mathrm{I}}+H_{\mathrm{B}}=H_{\mathrm{I}}^{(2)}+H_{\mathrm{B}}^{(2)}+H^{(n-1)}$ from Eq. (28), the initial PDF is simply a product PDF of the two-particle PDF and the reduced bath particle PDF. The reduced bath is in equilibrium with the inverse temperature β, but the two-particle system is not in equilibrium for $\beta'\neq\beta$. As the dynamics of the two-particle system and the reduced bath are completely decoupled, the reduced bath remains in equilibrium at later times, implying $\langle-\Delta\ln\rho\rangle=\langle-\Delta\ln\rho^{(2)}\rangle$ and $\langle\Delta H_{\mathrm{B}}^{(n-1)}\rangle=\langle\Delta H_{\mathrm{I}}^{(n-1)}\rangle=0$. Therefore, with this initial condition, the average heats are exactly given by $\langle Q_\alpha\rangle=\langle W\rangle-\langle\Delta\widehat{H}_\alpha^{(2)}\rangle$ with $\langle\Delta\widehat{H}_\alpha^{(n-1)}\rangle=0$.

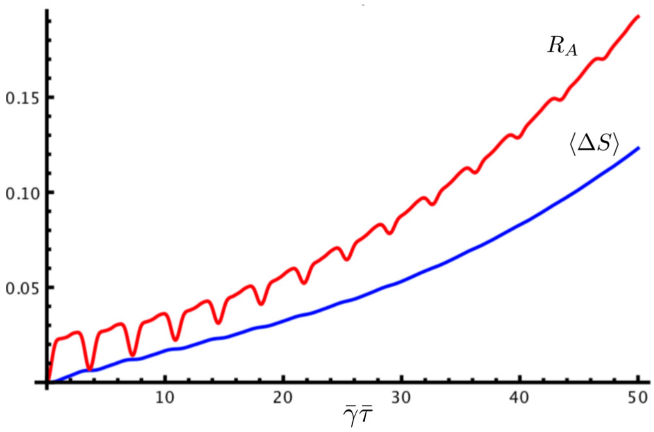

FIG. 1. R_A (red, upper line) and $\langle \Delta S\rangle$ (blue, lower line) in units of k_B vs $\bar{\gamma}\bar{\tau}$, with dimensionless parameters $\bar{\gamma} \equiv \gamma/\sqrt{mk}$ and $\bar{\tau} \equiv \tau/\sqrt{m/k}$, in the presence of the sliding harmonic potential with the dimensionless speed $\bar{u} \equiv u\sqrt{\beta m}$. We set $n = 2$, $\beta' = 1.5\beta$, $\bar{u} = 0.02$, $\bar{\gamma} = 30$, $\mu = m$, and $\kappa = k$. R_A is always non-negative and not smaller than $\langle \Delta S\rangle$. The difference between the two quantities is given by the relative entropy $\langle D_a(\tau)\rangle$. In contrast with $\langle \Delta S\rangle$ that always increases with time τ, R_A shows a nonmonotonic behavior in time. This nonmonotonicity is significant for small values of u and n. For large u and n, the positive heat production linearly proportional to time τ becomes dominant and the nonmonotonous behavior becomes negligible.

This PDF also satisfies the initial condition of type Eq. 11(a), which guarantees $R_\mathrm{A} \geqslant 0$ shown in Eq. (17). We can evaluate $\widetilde{H}_\mathrm{S}$ from Eq. (12), which turns out to be a constant independent of the system state in our simple model [47], and thus $\langle \Delta \widetilde{H}_\mathrm{S}\rangle = 0$ in Eq. (17). Thus, we can calculate the time evolution of R_A from the marginal PDF ρ_S in Eq. (39) and $\langle Q_\mathrm{S}\rangle = \langle W\rangle - \langle \Delta \widehat{H}_\mathrm{S}^{(2)}\rangle$ with Eqs. (41) and (47).

Finally, the initial condition for the two-particle system is deduced from Eq. (49) as $\rho^{(2)}(\vec{q}, 0) \propto \exp[-\beta' H_\mathrm{S}^{(2)} - \beta(H_\mathrm{B}^{(2)} + H_\mathrm{I}^{(2)})]$, which yield

$$\mathsf{A}_0 = \frac{\beta'}{\beta}\mathsf{A}_\mathrm{S} + \mathsf{A}_\mathrm{B} + \mathsf{A}_\mathrm{I}, \tag{50}$$

which will be used in the calculation of $\langle \Delta \widehat{H}_\alpha^{(2)}\rangle$ in Eq. (47). We also use the simplified expression for the average Shannon entropy change in the harmonic system such as

$$\begin{aligned}\langle -\Delta \ln \rho_\mathrm{S}\rangle &= -\Delta\left[\frac{1}{2}\ln\det\widehat{\mathsf{S}}_\tau - \frac{\beta}{2}\mathrm{Tr}\widehat{\mathsf{S}}_\tau\widehat{\mathsf{S}}_\tau^{-1}\right]\\ &= \frac{1}{2}\left[\ln\det\widehat{\mathsf{S}}_\tau^{-1} - \ln\det\widehat{\mathsf{S}}_0^{-1}\right],\end{aligned} \tag{51}$$

and similarly for $\langle -\Delta \ln \rho^{(2)}\rangle$.

Our numerical results are shown in Fig. 1, which is consistent with the expectation such that $R_\mathrm{A} \geqslant \langle \Delta S\rangle$ and R_A is not monotonous, in particular for weak nonequilibrium with small values of u and n. As u and n increase, the heat production is dominant in the entropy production, and thus the nonmonotonicity of R_S weakens.

For completeness, we report the explicit form of the average relative entropy $\langle D_\mathrm{a}\rangle = R_\mathrm{A} - \langle \Delta S\rangle$ as

$$\begin{aligned}\langle D_\mathrm{a}(\rho(\tau)||\widetilde{\rho}(\tau))\rangle = &\frac{1}{2}\ln\frac{\det\mathsf{A}_\tau}{\det\widehat{\mathsf{S}}_\tau}\frac{\det\widehat{\mathsf{S}}_0}{\det\mathsf{A}_0}\\ &+\frac{\beta}{2}\mathrm{Tr}\big(\mathsf{A}_\tau^{-1} - \mathsf{A}_0^{-1}\big)(\mathsf{A}_\mathrm{I} + \mathsf{A}_\mathrm{B})\\ &+\frac{\beta}{2}\vec{d}(\tau)\cdot(\mathsf{A}_\mathrm{I} + \mathsf{A}_\mathrm{B})\cdot\vec{d}(\tau).\end{aligned} \tag{52}$$

D. Three heat distributions

We now investigate fluctuations of the three heats in Eq. (43). The generating function for the heat distribution is given as

$$\begin{aligned}\mathcal{G}_\alpha(\lambda) &= \langle e^{-\beta\lambda Q_\alpha}\rangle = \langle e^{-\beta\lambda(W - \Delta\widehat{H}_\alpha)}\rangle\\ &= \big\langle e^{\beta\lambda(W - \Delta\widehat{H}_\alpha^{(2)})}\big\rangle_{(2)}\big\langle e^{\beta\lambda\Delta\widehat{H}_\alpha^{(n-1)}}\big\rangle_{(n-1)}\\ &\equiv \mathcal{G}_\alpha^{(2)}(\lambda, u)\cdot\mathcal{G}_\alpha^{(n-1)}(\lambda),\end{aligned} \tag{53}$$

where $\langle\cdot\rangle_{(2)}$ and $\langle\cdot\rangle_{(n-1)}$ are the average over two-particle state variables $\vec{q}$ and over reduced bath coordinates $\{x_i', p_i'\}$ for $t = [0, \tau]$, respectively. The u dependence comes from the two-particle system Hamiltonian in Eq. (26).

With the initial condition taken in Eq. (49) ($\beta' = \beta$), the reduced noninteracting bath particles are always in equilibrium, independent from the evolution of the two-particle system. Thus, the generating function for the reduced system should be factorized as

$$\mathcal{G}_\alpha^{(n-1)}(\lambda) = g_\alpha(\lambda)^{n-1}, \tag{54}$$

where

$$g_\alpha(\lambda) = \big\langle e^{\beta\lambda\Delta\widehat{H}_\alpha^i}\big\rangle, \tag{55}$$

with

$$\widehat{H}_\mathrm{S}^i = 0\,,\ \widehat{H}_\mathrm{B}^i = \frac{\kappa}{2}x_i'^2\,,\ \widehat{H}_\mathrm{SB}^i = \frac{\kappa}{2}x_i'^2 + \frac{1}{2m}p_i'^2. \tag{56}$$

Note that for a Langevin system with no explicit time dependence in Hamiltonian $H(\tau) = H(q(\tau))$ for state q,

$$\begin{aligned}&\langle e^{\beta\lambda H(\tau) - \beta\lambda H(0)}\rangle\\ &\quad= \int dq_\tau\int dq_0\int D[q(t)]e^{\beta\lambda H(\tau)}\,\Pi[q(t)]\,\frac{e^{-\beta(1+\lambda)H(0)}}{Z(\beta)}\\ &\quad\longrightarrow \frac{Z(\beta(1+\lambda))}{Z(\beta)}\frac{Z(\beta(1-\lambda))}{Z(\beta)}\quad\text{as } \tau\to\infty,\end{aligned} \tag{57}$$

where $\Pi[q(t)]$ is the path probability, Z is the partition function, and it is used that the final PDF becomes Boltzmann as $\tau \to \infty$, independent of an initial PDF. Using this formula, we find

$$\mathcal{G}_\alpha^{(n-1)}(\lambda) = \frac{1}{(1-\lambda^2)^{\frac{1}{2}d_\alpha^{(n-1)}}}, \tag{58}$$

where $d_\alpha^{(n-1)}$ is the number of the degrees of freedom in $\widehat{H}_\alpha^{(n-1)}$, which is equal to 0, $n-1$, $2(n-1)$ for $\alpha =$ S, B, SB, respectively.

The most difficult task is to compute $\mathcal{G}_\alpha^{(2)}(\lambda, u)$, for which we can extend the results from the recent study on a single-particle Langevin system in the underdamped case under the same moving potential [43]. In Eq. (53), $\beta\lambda\widehat{H}_\alpha^{(2)}$ in the exponent effectively alters the PDFs at $t=0$ and $t=\tau$. Note that $\widehat{H}_\alpha^{(2)}$ for the two-particle system is a function of $\vec{r}=\vec{q}-\vec{u}t$, as can be seen in Eq. (36). For $\vec{r}=\vec{z}+\vec{d}(t)$, the path integral is to be performed over stochastic variables $\vec{z}(t)$. We can separate $H^{(2)}$ in two parts, i.e., z-dependent and -independent parts, as

$$\begin{aligned}\widehat{H}_\alpha^{(2)}(\vec{r}(t)) &= \tfrac{1}{2}\vec{d}(t)\cdot\widehat{\mathsf{A}}_\alpha\cdot\vec{d}(t)\\ &\quad+\vec{d}(t)\cdot\widehat{\mathsf{A}}_\alpha\cdot\vec{z}(t)+\tfrac{1}{2}\vec{z}(t)\cdot\widehat{\mathsf{A}}_\alpha\cdot\vec{z}(t)\\ &= f_\alpha(t)+\vec{d}(t)\cdot\widehat{\mathsf{A}}_\alpha\cdot\vec{z}(t)+H_\alpha^{(2)}(\vec{z}(t)),\end{aligned}\tag{59}$$

where $\widehat{\mathsf{A}}_\alpha=\mathsf{A}_\mathrm{S},\ \mathsf{A}_\mathrm{S}+\mathsf{A}_\mathrm{I},\ \mathsf{A}_\mathrm{eq}$ for $\alpha=\mathrm{S},\mathrm{B},\mathrm{SB}$, respectively, consistent with the definition of $\widehat{H}_\alpha^{(2)}$ in terms of H_α. Then, using the formula in Eq. (57), we can rewrite the two-particle generating function in the long-time limit as

$$\begin{aligned}\mathcal{G}_\alpha^{(2)}(\lambda,u) &= e^{\beta\lambda h_\alpha(\tau)}\frac{Z_\alpha(\beta,\lambda)}{Z(\beta)}\frac{Z_\alpha(\beta,-\lambda)}{Z(\beta)}\\ &\quad\times\left\langle\exp\left[\vec{d}(\tau)\cdot\widehat{\mathsf{A}}\cdot\vec{z}_\tau+\beta\lambda ku\int_0^\tau dt\,z_x(t)\right]\right\rangle_{(2)}^{\mathrm{ren}},\end{aligned}\tag{60}$$

where

$$h_\alpha(\tau)=f_\alpha(\tau)+ku\int_0^\tau dt d_x(t).$$

Here, the superscript "ren" means the renormalized path integral due to the alteration of the initial distribution and the constraint on the final distribution, written as

$$\begin{aligned}\langle\mathcal{F}[\vec{z}(t)]\rangle_{(2)}^{\mathrm{ren}} &= \int d\vec{z}_\tau\int d\vec{z}_0\int D[\vec{z}(t)]\frac{Z(\beta)e^{\beta\lambda\widehat{H}_\alpha^{(2)}(\vec{z}_\tau)}}{Z_\alpha(\beta,-\lambda)}\\ &\quad\times\Pi[\vec{z}(t)]\,\mathcal{F}[\vec{z}(t)]\,\frac{e^{-\beta[H^{(2)}(\vec{z}_0)+\lambda\widehat{H}_\alpha^{(2)}(\vec{z}_0)]}}{Z_\alpha(\beta,\lambda)}.\end{aligned}\tag{61}$$

Here, $Z_\alpha(\beta,\pm\lambda)$ $[Z(\beta)=Z_\alpha(\beta,0)]$ is the partition function due to an altered PDF for $t=0$ $(+\lambda)$ or $t=\tau$ $(-\lambda)$, given as

$$Z_\alpha(\beta,\lambda)=\int d\vec{z}\,e^{-\beta[H^{(2)}(\vec{z})+\lambda\widehat{H}_\alpha^{(2)}(\vec{z})]}=\sqrt{\frac{(2\pi/\beta)^4}{\det(\mathsf{A}_\mathrm{eq}+\lambda\widehat{\mathsf{A}}_\alpha)}}.\tag{62}$$

In Eq. (60), the renormalized path integral $I_\alpha(\lambda,u)$ is to be performed for the exponential of linear terms in $\vec{z}(t)$ and can be carried out by using the cumulant expansion. As a result, we find

$$\begin{aligned}\ln I_\alpha(\lambda,u) &= \frac{1}{2}\vec{d}(\tau)\cdot\widehat{\mathsf{A}}_\alpha\tilde{\mathsf{C}}_\alpha(\tau,\tau)\widehat{\mathsf{A}}_\alpha\cdot\vec{d}(\tau)\\ &\quad+\beta\lambda ku\int_0^\tau dt[\vec{d}(\tau)\cdot\widehat{\mathsf{A}}_\alpha\tilde{\mathsf{C}}_\alpha(\tau,t)]_1\\ &\quad+\frac{(\beta\lambda ku)^2}{2}\int_0^\tau dt\int_0^\tau dt'[\tilde{\mathsf{C}}_\alpha(t,t')]_{11},\end{aligned}\tag{63}$$

where $[\cdot]_1$, $[\cdot]_{11}$ denote the first component of a vector and (1,1) element of a matrix. In this equation, $\tilde{\mathsf{C}}_\alpha(t,t')$ is the renormalized correlation function (matrix) obtained from

$$\begin{aligned}&\langle e^{\vec{l}(t)\cdot\vec{z}(t)+\vec{l}(t')\cdot\vec{z}(t')}\rangle_{(2)}^{\mathrm{ren}}\\ &\quad=e^{\frac{1}{2}\vec{l}(t)\cdot\tilde{\mathsf{C}}_\alpha(t,t)\cdot\vec{l}(t)}e^{\frac{1}{2}[\vec{l}(t')\cdot\tilde{\mathsf{C}}_\alpha(t,t)\cdot\vec{l}(t')+\vec{l}(t)\cdot\tilde{\mathsf{C}}_\alpha(t,t')\cdot\vec{l}(t')+\vec{l}(t')\cdot\tilde{\mathsf{C}}_\alpha(t',t)\cdot\vec{l}(t)]},\end{aligned}$$

where the α dependence comes from the alteration of PDFs due to $\widehat{H}_\alpha^{(2)}$. Then, we can find, for $t>t'$,

$$\begin{aligned}\tilde{\mathsf{C}}_\alpha(t,t') &= \beta^{-1}\big[e^{-\mathsf{F}(t-t')}\mathsf{A}_\mathrm{eq}^{-1}-\lambda\mathsf{B}_\alpha(t,t')\big]\\ &\quad+\beta^{-1}\lambda\big[\mathsf{A}_\mathrm{eq}^{-1}-\mathsf{B}_\alpha(t,t)\big]\\ &\quad\times\mathsf{D}_\alpha(t,t')\big[\mathsf{A}_\mathrm{eq}^{-1}-\mathsf{B}_\alpha(t',t')\big],\end{aligned}\tag{64}$$

with

$$\begin{aligned}\mathsf{B}_\alpha(t,t') &= e^{-\mathsf{F}t}(\mathsf{A}_\mathrm{eq}+\lambda\widehat{\mathsf{A}}_\alpha)^{-1}\widehat{\mathsf{A}}_\alpha\mathsf{A}_\mathrm{eq}^{-1}e^{-\mathsf{F}^\mathrm{t}t'},\\ \mathsf{D}_\alpha(t,t') &= e^{-\mathsf{F}^\mathrm{t}(\tau-t)}\mathsf{A}_\mathrm{eq}(\mathsf{A}_\mathrm{eq}-\lambda\widehat{\mathsf{A}}_\alpha)^{-1}\widehat{\mathsf{A}}_\alpha e^{-\mathsf{F}^\mathrm{t}(\tau-t')},\end{aligned}$$

where it becomes the original correlation function for $\lambda=0$.

After some algebra, Eq. (60) can be computed using Eqs. (62)–(64) in the long τ limit, where we neglect terms having $e^{-\mathsf{F}\tau}$ and $e^{-\mathsf{F}^\mathrm{t}\tau}$. As a result, the generating function for the two-particle system is given as

$$\mathcal{G}_\alpha^{(2)}(\lambda,u)\simeq\frac{e^{-n\tau w\lambda(1-\lambda)-nwb_\alpha\lambda^3/[2(1+\lambda)]}}{(1-\lambda^2)^{\frac{1}{2}d_\alpha^{(2)}}},\tag{65}$$

for $d_\alpha^{(2)}$ is the number of degrees of freedom in $\widehat{H}_\alpha^{(2)}$, where we defined the work rate per bath particles as $w=\beta\langle\dot{W}\rangle/n=\beta\gamma u^2$ for $\tau\to\infty$ from Eq. (42). Noting $\widehat{H}_\alpha^{(2)}=H_\mathrm{S}^{(2)},\ H_\mathrm{S}^{(2)}+H_\mathrm{I}^{(2)},\ H^{(2)}$, then $d_\alpha^{(2)}=2,\ 3,\ 4$ for $\alpha=\mathrm{S}$, B, and SB, respectively. For simplicity, we only show terms of order up to $\mathcal{O}(\tau)$, keeping divergent terms for $\lambda=\pm1$. We find the coefficient for divergence in the exponent as

$$b_\alpha=\begin{cases}\frac{n\gamma}{k}+\frac{\mu}{n\gamma} & \text{for } \alpha=\mathrm{S}\\ \frac{(k+n\kappa)\gamma}{\kappa k}+\frac{\mu}{n\gamma} & \text{for } \alpha=\mathrm{B}\\ \frac{(k+n\kappa)\gamma}{\kappa k}+\frac{\mu+nm}{n\gamma} & \text{for } \alpha=\mathrm{SB}.\end{cases}\tag{66}$$

Finally, the generating function can be obtained as $\mathcal{G}_\alpha^{(2)}(\lambda,u)\mathcal{G}_\alpha^{(n-1)}(\lambda)$ with Eq. (58).

From the inverse-Fourier transformation of $\mathcal{G}_\alpha(\lambda)$, we have the integral representation of the heat distribution function for dimensionless heat rate q defined as $\beta Q=\beta q\langle W\rangle|_{\tau\to\infty}=n\tau wq$ as

$$\mathcal{P}_\alpha(q)=\int_{-i\infty}^{i\infty}\frac{d\lambda}{2\pi i}\frac{e^{-n\tau w[\lambda(1-\lambda)-q\lambda]-nwb_\alpha\lambda^3/[2(1+\lambda)]}}{(1-\lambda^2)^{\nu_\alpha}},\tag{67}$$

where $\nu_\alpha=(d_\alpha^{(n-1)}+d_\alpha^{(2)})/2$ is given as

$$\nu_\alpha=\begin{cases}1 & \text{for } \alpha=\mathrm{S}\\ 1+\frac{n}{2} & \text{for } \alpha=\mathrm{B}\\ 1+n & \text{for } \alpha=\mathrm{SB}.\end{cases}\tag{68}$$

For the equilibrium case with $u=0$ $(w=0)$, the heat distribution function can be exactly obtained. Starting from equilibrium, the total system remains in equilibrium so that there is no average heat production. However, fluctuations still occur, even in equilibrium. In this case, the above q is not a relevant variable. We define a dimensionless heat as $r=\beta Q$.

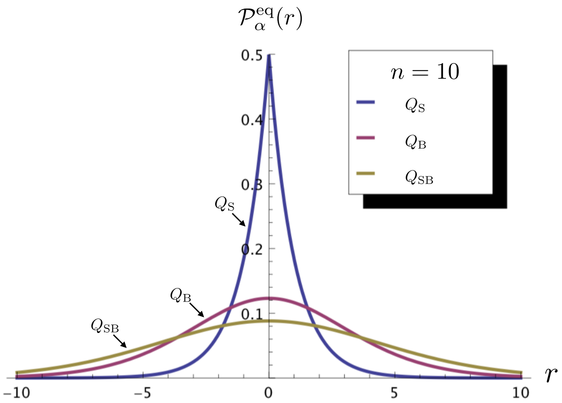

FIG. 2. Plots for $\mathcal{P}^{\rm eq}_{\alpha}(r)$ for $r = \beta Q$ (heat in units of β^{-1}). The distributions become broader for larger n. For the same n, $\mathcal{P}^{\rm eq}_{\rm SB}$ is the broadest and $\mathcal{P}^{\rm eq}_{\rm S}$ is the sharpest. $\mathcal{P}^{\rm eq}_{\rm S}(r) = e^{-|r|}/2$ is independent of n.

For the long-time (τ) limit, the equilibrium heat distribution function for r can be found as

$$\mathcal{P}^{\rm eq}_{\alpha}(r) = \int_{-\infty}^{\infty}\frac{d\lambda}{2\pi}\frac{e^{ir\lambda}}{(1+\lambda^2)^{\nu_\alpha}} = \frac{(|r|/2)^{\nu_\alpha-1/2}}{\sqrt{\pi}\,\Gamma(\nu_\alpha)}K_{\nu_\alpha-1/2}(|r|), \tag{69}$$

where $K_\mu(z)$ is the modified Bessel function of the second kind, defined for $z > 0$ as

$$K_\mu(z) = \frac{\Gamma(\mu+1/2)(2z)^\mu}{\sqrt{\pi}}\int_0^\infty dt\,\frac{\cos t}{(t^2+z^2)^{\mu+1/2}}. \tag{70}$$

Figure 2 shows a clear difference in three heat distributions in the equilibrium state depending on n, having zero mean as expected. $\mathcal{P}^{\rm eq}_{\rm S}(r) = e^{-|r|}/2$ is independent of n, which can be understood because the system effectively interacts with a single representative particle at the center-of-mass position. We remark that it is equal to that for the single-particle Langevin system [43], which supports the physical relevance of our finite-number-bath model in the long-time behavior. The broadness of $\mathcal{P}^{\rm eq}_{\rm B}(r)$ with larger n can be simply understood because the observable r is the total heat for the bath including noninteracting n reduced bath particles.

It is interesting to note that the three heat PDFs are quite different even in the long-time limit and, furthermore, do not depend on any parameter, including the coupling constant κ characterizing the strength of the interaction Hamiltonian $H_{\rm I}$. It implies that the difference between the three PDFs does not go away even in the conventional weak-coupling limit ($\kappa \to 0$). A very slow relaxation mode is induced by vanishing κ so that the energy transfer between the system and bath becomes extremely slower than the dynamics of the system variable (x, p), which is not desirable in the conventional weak-coupling models. Therefore, our simple model may not describe realistic situations for a very small κ.

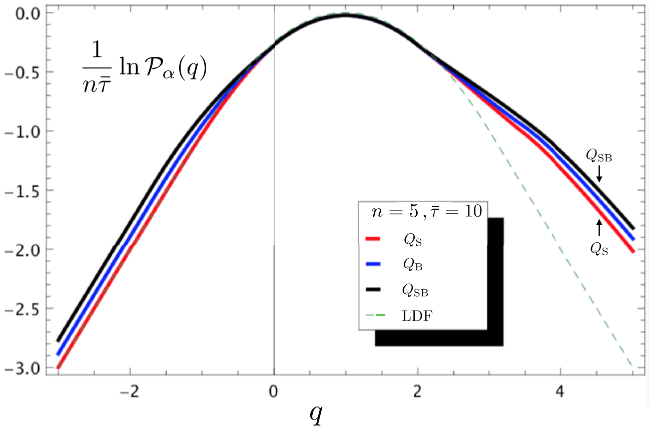

FIG. 3. Plots of $\ln\mathcal{P}_\alpha(q)/(n\bar{\tau})$ vs q by interpolating the curves in three piecewise regions. The dimensionless parameters are $\bar{\tau} = \tau/\sqrt{m/k}$, $q = \beta Q/(n\bar{w}\bar{\tau})$, $\bar{w} = \bar{\gamma}\bar{u}^2$ in the long-time limit. We plot the curves for the parameter values of $\bar{u} = 1$, $\bar{\gamma} = 1$, $\mu = m$, and $\kappa = k$. The dashed line is the n^{-1} large deviation function.

For the nonequilibrium case, we evaluate the integral in Eq. (67) in the long-time limit. The difference between the three heat PDFs is encoded in b_α and ν_α. The leading order result is independent of the details of power-law singularities associated with ν_α and b_α, which yields the large deviation function (LDF) as

$$\ln\mathcal{P}_\alpha(q) \sim \max_\lambda n\tau w[q\lambda - f(\lambda)], \tag{71}$$

where $f(\lambda) = \lambda(1-\lambda)$ for $-1 < \lambda < 1$, and ∞ otherwise. Thus, the three PDFs in nonequilibrium situations are equivalent up to the leading order, where the nonzero heat rate is dominant. This feature is very different from the equilibrium case where the average heat rate is zero. In order to identify the difference in the three PDFs, we should calculate the next order, which arises from the singularities around $\lambda = \pm 1$. We use the *modified* saddle-point approximation developed to deal with singularities in the saddle-point integration [42,48,49].

The saddle point λ^* is found from the extremum condition for the logarithm of the integrand of Eq. (67), given as

$$0 = -2\lambda^* + 1 - q - \frac{b_\alpha\lambda^{*3}}{2\tau(1+\lambda^*)^2} - \frac{2\nu_\alpha\lambda^*}{n\tau w(1-\lambda^{*2})}, \tag{72}$$

where the most divergent terms are kept near $\lambda = \pm 1$. The saddle point λ^* occurs for $-1 < \lambda^* < 1$. We find the heat distribution functions for three piecewise regions in q as follows:

(1) For λ^* far from ± 1 and $-1 < \lambda^* < 1$, the singular terms at $\lambda^* = \pm 1$ in Eq. (72) can be neglected, so $\lambda^* = (1-q)/2$ and $-1 < q < 3$. Equation (67) can be carried out by the regular saddle-point approximation for large τ.

(2) For $\lambda^* \simeq 1$, we define $\delta\lambda_+ = \lambda^* - 1 < 0$. Then, the saddle-point condition leads to

$$\delta\lambda_+ \simeq \frac{1}{4}\left[-(q+1) - \sqrt{(q+1)^2 + \frac{8\nu_\alpha}{n\tau w}}\right], \tag{73}$$

for which there are two cases: $(1+q)^2 \gg \frac{8\nu_\alpha}{n\tau w}$ and $(1+q)^2 \ll \frac{8\nu_\alpha}{n\tau w}$. The latter is restricted to a very narrow region, so we only consider the former case with

$$\delta\lambda_+ \simeq -\frac{4\nu_\alpha}{n\tau w|q+1|}, \tag{74}$$

for $q < -1$ and $|q+1| \gg 1$. The usual saddle-point approximation is not useful because $(1-\lambda^2)^{-\nu_\alpha}$ in the integrand cannot be expanded about $\lambda^* \simeq 1$. We use a new variable $v = 1 + (\lambda - \lambda^*)/\delta\lambda_+$, so that $(1-\lambda)^{-1} = -(\delta\lambda_+ v)^{-1}$ which is treated exactly. Then, we get

$$\mathcal{P}_\alpha(q) \simeq \frac{e^{\tau n w q}}{|\delta\lambda_+|^{\nu_\alpha - 1}(-2^{\nu_\alpha})} \int_{1-i\infty}^{1+i\infty} \frac{dv}{2\pi i} \frac{e^{4\nu_\alpha v}}{v^{\nu_\alpha}}, \tag{75}$$

where the multiplicative factor leads to the logarithmic correction of $\mathcal{O}(\ln\tau)$ to the LDF and the remaining integral is of $\mathcal{O}(1)$, which is negligible.

(3) For $\lambda^* \simeq -1$, we define $\delta\lambda_- = \lambda^* + 1 > 0$, and the saddle-point condition leads to

$$2\delta\lambda_- - (q-3) + \frac{b_\alpha}{2\tau(\delta\lambda_-)^2} = 0. \tag{76}$$

We omit the narrow region for $q - 3 \ll \delta\lambda_-$. For $q - 3 \gg \delta\lambda_-$, we get

$$\delta\lambda_- = \left[\frac{b_\alpha}{2\tau(q-3)}\right]^{1/2} \tag{77}$$

for $q - 3 \gg [b_\alpha/(2\tau)]^{1/3}$. Similarly to the case for $q < -1$, we introduce variable $u = 1 + (\lambda - \lambda^*)/\delta\lambda_-$, so that $(1+\lambda)^{-1} = (\delta\lambda_- u)^{-1}$ using a new variable u. Then, we find

$$\mathcal{P}_\alpha(q) \simeq \frac{e^{-\tau n w(p-2)}}{(\delta\lambda_-)^{\nu_\alpha - 1} 2^{\nu_\alpha}} \int_{1-i\infty}^{1+i\infty} \frac{du}{2\pi i} \frac{e^{nw\sqrt{\tau(q-3)/2}(u+1/u)}}{u^{\nu_\alpha}}. \tag{78}$$

The integral over u can be found by using the standard saddle-point approximation for the saddle point $u^* = 1$ given from $d(u + 1/u)/du|_{u=u^*} = 0$. It gives rise to the correction of $\mathcal{O}(\tau^{1/2})$ to the LDF.

We finally find the heat distribution functions for three piecewise regions: (1) far from $\lambda^* = \pm 1$ corresponding to $-1 < q < 3$ (center); (2) $\lambda^* \simeq 1$ corresponding to $q < -1$ (left wing); (3) $\lambda^* \simeq -1$ corresponding to $q > 3$ (right wing), given as

$$\mathcal{P}_\alpha(q) = \begin{cases} \exp\left[-\frac{n\tau}{4} w(1-q)^2 - \frac{1}{2}\ln(w\tau)\right] & ; -1 < q < 3 \\ \exp[n\tau w q + (\nu_\alpha - 1)\ln(w\tau)] & ; q < -1,\ |q+1| \gg \left(\frac{nw\tau}{8\nu_\alpha}\right)^{-1/2} \\ \exp\left[-n\tau w(q-2) + n\sqrt{2w^2 b_\alpha \tau(q-3)} + \frac{1}{2}\left(\nu_\alpha - \frac{3}{2}\right)\ln[w^2 b_\alpha \tau(q-3)]\right] & ; q - 3 \gg \left(\frac{2\tau}{b_\alpha}\right)^{-1/3}. \end{cases} \tag{79}$$

The LDF is found in the leading order as shown in the terms proportional to τ in the exponent, which is consistent with Eq. (71). The corrections to the LDF manifest the difference in the three heat distributions. Figure 3 shows the three heat distributions, for which the curves in three piecewise regions are interpolated for the purpose of guidance. The LDF is drawn by a broken curve. The most significant corrections to the LDF appear in the right wing for $q > 3$, which originates from the so-called everlasting initial memory of having an excessive value for $\widehat{H}_\alpha$ in Eq. (5), which was reported as a characteristic of nonequilibrium heats in previous works [42,48–51].

We find in Fig. 3 or Eq. (79) that $\mathcal{P}_\mathrm{B}(q)$ becomes slightly flatter for smaller κ, due to increasing b_B in Eq. (66). This can be understood as an influence from increasing distance fluctuations of bath particles around the system due to the weak interaction (small κ). On the other hand, $\mathcal{P}_\mathrm{S}(q)$ for the system trapped by the harmonic potential (finite k) remains unchanged. A clearer difference is expected for the heat distribution functions for finite τ. We note that the $\kappa \to 0$ limit is singular because b_α diverges, which also indicates that our model does not show the proper weak-coupling regime in the simple $\kappa \to 0$ limit.

VI. SUMMARY

We find the three different heats in system and bath flowing in different directions, which were conventionally regarded as the same quantity flowing from one place to another. We show clear differences between them by rigorously treating the interaction Hamiltonian, especially shown in $\Delta H_\mathrm{I} = Q_\mathrm{S} - Q_\mathrm{B}$. These differences can be regarded as an indication of a strong coupling between the system and bath. The three different heats are manifested in the various fluctuation theorems and distinguish themselves in their fluctuations even with the same average value in the long-time limit. The theoretical expectations are explicitly confirmed from the calculation for the typical example with a sliding harmonic potential.

The weak-coupling regime is usually believed to be reached by a perturbation approach for small H_I, for which one might expect $\mathcal{P}_\mathrm{S}(q) \simeq \mathcal{P}_\mathrm{B}(q)$ in the extreme limit. However, it is not the case, at least for our example. In both the equilibrium and nonequilibrium cases, the three heat PDFs remains different even in the extreme $\kappa \to 0$ limit. We suspect that this simple limit by itself does not guarantee the standard weak-coupling limit in general theoretical models, and additional assumptions such as timescale separations should be incorporated.

The appearance of the three different heats or the two different heats for the nonthermostatted bath seems to be inevitable as we handle the interaction Hamiltonian rigorously for the strong-coupling case. It would be interesting to extend our classical study on various heats to quantum systems.

ACKNOWLEDGMENTS

This work was supported by the NRF Grants No. 2016R1D1A1A09918020 (C.K.), No. 2017R1D1A1B03030872 (J.U.), No. 2017R1D1A09000527 (J.Y.), and No. 2017R1D1A1B06035497 (H.P.), Korea.

[1] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993).

[2] D. J. Evans and D. J. Searles, Phys. Rev. E **50**, 1645 (1994).

[3] G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995); J. Stat. Phys. **80**, 931 (1995).

[4] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[5] C. Jarzynski, Phys. Rev. E **56**, 5018 (1997).

[6] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).

[7] J. Kurchan, J. Phys. A **31**, 3719 (1998).

[8] J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).

[9] T. Speck and U. Seifert, J. Phys. A **38**, L581 (2005); U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).

[10] M. Esposito and C. Van den Broeck, Phys. Rev. Lett. **104**, 090601 (2010).

[11] P. Talkner, E. Lutz, and P. Hänggi, Phys. Rev. E **75**, 050102(R) (2007).

[12] M. Campisi, P. Talkner, and P. Hänggi, Phys. Rev. Lett. **102**, 210401 (2009).

[13] M. Campisi, P. Hänggi, and P. Talkner, Rev. Mod. Phys. **83**, 771 (2011).

[14] P. Hänggi and P. Talkner, Nat. Phys. **11**, 108 (2015).

[15] J. Schnakenberg, Rev. Mod. Phys. **48**, 571 (1976).

[16] K. Sekimoto, J. Phys. Soc. Jpn. **68**, 1448 (1999).

[17] T. M. Nieuwenhuizen and A. E. Allahverdyan, Phys. Rev. E **66**, 036102 (2002).

[18] C. Hörhammer and H. Büttner, J Stat. Phys. **133**, 1161 (2008).

[19] I. Kim and G. Mahler, Phys. Rev. E **81**, 011101 (2010).

[20] M. Esposito, K. Lindenberg, and C. Van den Broeck, New J. Phys. **12**, 013013 (2010).

[21] L. Pucci, M. Esposito, and L. Peliti, J. Stat. Mech. (2013) P04005.

[22] J. M. Horowitz and J. M. R. Parrondo, New J. Phys. **15**, 085028 (2013).

[23] F. W. J. Hekking and J. P. Pekola, Phys. Rev. Lett. **111**, 093602 (2013).

[24] J. M. Horowitz and T. Sagawa, J. Stat. Phys. **156**, 55 (2014).

[25] M. Silaev, T. T. Heikkilä, and P. Virtanen, Phy. Rev. E **90**, 022103 (2014).

[26] J. Ankerhold and J. P. Pekola, Phys. Rev. B **90**, 075421 (2014).

[27] R. Gallego, A. Riera, and J. Eisert, New J. Phys. **16**, 125009 (2015).

[28] M. Carrega, P. Solinas, A. Braggioi, M, Sassetti, and U. Weiss, New J. Phys. **17**, 045030 (2015).

[29] M. Esposito, M. A. Ochoa, and M. Galperin, Phys. Rev. Lett. **114**, 080602 (2015).

[30] U. Seifert, Phys. Rev. Lett. **116**, 020601 (2016).

[31] P. Strasberg and M. Esposito, Phys. Rev. E **95**, 062101 (2017).

[32] E. Iyoda, K. Kaneko, and T. Sagawa, Phys. Rev. Lett. **119**, 100601 (2017).

[33] K. Funo and H. T. Quan, Phys. Rev. Lett. **121**, 040602 (2018).

[34] H. Park, J. Kor. Phys. Soc. **72**, 1413 (2018).

[35] As there is no membership change of bath particles in interactions, there would be considerable memory remaining, which might cause some unphysical results. However, the qualitative behaviors of the three heat PDFs seem not seriously damaged.

[36] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. **98**, 3658 (2001).

[37] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **89**, 050601 (2002).

[38] G. M. Wang, J. C. Reid, D. M. Carberry, D. R. M. Williams, E. M. Sevick, and D. J. Evans, Phys. Rev. E **71**, 046142 (2005).

[39] R. van Zon and E. G. D. Cohen, Phys. Rev. E **67**, 046102 (2003).

[40] R. van Zon and E. G. D. Cohen, Phys. Rev. Lett. **91**, 110601 (2003).

[41] A. Imparato, L. Peliti, G. Pesce, G. Rusciano, and A. Sasso, Phys. Rev. E **76**, 050101(R) (2007).

[42] K. Kim, C. Kwon, and H. Park, Phys. Rev. E **90**, 032117 (2014).

[43] C. Kwon, J. Kor. Phys. Soc. **73**, 866 (2018).

[44] C. Gardiner, *Stochastic Methods: A Handbook Handbook for the Natural and Social Sciences*, 4th ed. (Springer, New York, 2010).

[45] C. Kwon, P. Ao, and D. J. Thouless, Proc. Nat. Acad. Sci. USA **102**, 13029 (2005).

[46] C. Kwon, J. D. Noh, and H. Park, Phys. Rev. E **83**, 061145 (2011).

[47] In our simple model without the bath potential ($U_\mathrm{B} = 0$), in fact one cannot evaluate the bath partition function Z_B without a confined volume. We introduce a confined volume and find that $\widetilde{H}_\mathrm{S}$ becomes a constant without any dependence of the system variable.

[48] J. S. Lee, C. Kwon, and H. Park, Phys. Rev. E **87**, 020104(R) (2013).

[49] J. S. Lee, C. Kwon, and H. Park, J. Stat. Mech. (2013) P11002.

[50] J. Farago, J. Stat. Phys. **107**, 781 (2002); Physica A (Amsterdam) **331**, 69 (2004).

[51] A. Puglisi, L. Rondoni, and A. Vulpiani, J. Stat. Mech. (2006) P08010.