Strong coupling thermodynamics and stochastic thermodynamics from the unifying perspective of time-scale separation

Mingnan Ding,[1] Zhanchun Tu,[2] and Xiangjun Xing[1,3,4,*]

[1]*Wilczek Quantum Center, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China*
[2]*Department of Physics, Beijing Normal University, Beijing 100875, China*
[3]*Tsung-Dao Lee Institute, Shanghai Jiao Tong University, Shanghai 200240, China*
[4]*Shanghai Research Center for Quantum Sciences, Shanghai 201315, China*

(Received 26 October 2020; revised 26 May 2021; accepted 23 November 2021; published 7 January 2022)

Assuming time-scale separation, a simple and unified theory of thermodynamics and stochastic thermodynamics is constructed for small classical systems strongly interacting with their environments in a controllable fashion. The total Hamiltonian is decomposed into a bath part and a system part, the latter being the *Hamiltonian of mean force*. Both the conditional equilibrium of the bath and the reduced equilibrium of the system are described by canonical ensemble theories with respect to their own Hamiltonians. The bath free energy is independent of the system variables and the control parameter. Furthermore, the weak coupling theory of stochastic thermodynamics becomes applicable *almost verbatim*, even if the interaction and correlation between the system and its environment are strong and varied externally. We further discuss a simple scenario where the present theory fits better with the common intuition about system entropy and heat.

DOI: 10.1103/PhysRevResearch.4.013015

I. INTRODUCTION

One of the most significant discoveries of statistical physics in the past few decades is that thermodynamic variables can be defined on the level of the dynamic trajectory [1–3]. Studies of these fluctuating quantities in nonequilibrium processes have led to significant results such as fluctuation theorems [2] and the Jarzynski equality [3], as well as a much deeper understanding of the second law of thermodynamics.

Consider, for example, a small classical system with Hamiltonian $H(\boldsymbol{x}, \lambda)$ weakly interacting with its bath, such that the interaction energy and statistical correlation between the system and the bath are negligibly small. Here, $\boldsymbol{x} = (\boldsymbol{q}, \boldsymbol{p})$ are the canonical variables, and λ is an external control parameter. The differential work and heat at trajectory level are defined as

$$đ\mathscr{W} \equiv H(\boldsymbol{x}, \lambda + d\lambda) - H(\boldsymbol{x}, \lambda) \equiv d_\lambda H(\boldsymbol{x}, \lambda), \quad (1.1a)$$

$$đ\mathcal{Q} \equiv H(\boldsymbol{x} + d\boldsymbol{x}, \lambda) - H(\boldsymbol{x}, \lambda) \equiv d_{\boldsymbol{x}} H(\boldsymbol{x}, \lambda), \quad (1.1b)$$

respectively. Throughout this paper, we use the notations $d_\lambda H(\boldsymbol{x}, \lambda)$ and $d_{\boldsymbol{x}} H(\boldsymbol{x}, \lambda)$ for differentials of $H(\boldsymbol{x}, \lambda)$ due to variations of λ and of $\boldsymbol{x}$, respectively [4]. These notations will greatly simplify the presentation of our theory. With $H(\boldsymbol{x}, \lambda)$ identified as the fluctuating internal energy, the first law at the trajectory level then follows directly: $dH = d_\lambda H + d_{\boldsymbol{x}} H = đ\mathscr{W} + đ\mathcal{Q}$. Further using the time-reversal symmetry of Hamiltonian dynamics or Langevin dynamics, one can derive the Crooks function theorem, Jarzynski equality, and Clausius inequality. Mathematical expressions for various thermodynamic variables of weak coupling stochastic thermodynamics are shown in the center column of Table I of Sec. V. For pedagogical reviews, see, e.g., Refs. [2,3].

In recent years, there has been significant interest in generalizing thermodynamics and stochastic thermodynamics to small systems that are strongly coupled to the environment, both classical [5–16] and quantum [6,8,17–28]. Strong interactions between a system and its environment cause ambiguities in the definitions of system thermodynamic quantities [6,8]. If the system size is large and the interactions are short ranged, the correlations between system and bath are confined to the interfacial regions and hence do not influence the bulk properties of the system. This is indeed the reason why classical thermodynamics and statistical mechanics are so successful in describing the equilibrium properties of macroscopic systems, even if these systems may be strongly interacting with the environment near the interfaces. Small systems, however, have no "bulk," and their thermodynamic properties may be overwhelmingly dominated by their interactions and correlations with the environment. Should one relegate the interaction energy to the system or to the bath? Should one treat the mutual information between the system and bath variables as part of the system entropy or the bath entropy? There seems to be no general principle in favor of any particular answer. For critical and insightful discussions of these fundamental issues, see the recent articles by Jarzynski [7] and by Talkner and Hänggi [8].

*xxing@sjtu.edu.cn

Numerous versions [5–7,12,29] of strong coupling thermodynamic theories have been proposed in recent years. The theory that is probably the most influential was developed by Seifert [5] and critically evaluated by Talkner and Hänggi [6,8]. In this theory, one uses the *Hamiltonian of mean force* (HMF) $H_{\mathbf{X}}$ [16,30,31] to construct the equilibrium free energy $F = -T \ln \int e^{-\beta H_{\mathbf{X}}}$ and then *defines* equilibrium system energy and entropy via $E = \partial \beta F/\partial \beta$, $S = -\beta^2 \partial F/\partial \beta$. Whereas these relations exactly hold in equilibrium thermodynamics, they must be deemed as *definitions* of energy and entropy in Seifert's theory of strong coupling thermodynamics. Interestingly, these definitions correspond to the particular decomposition of total thermodynamic variables $A_{\text{tot}} = A_{\text{sys}} + A_{\text{bath}}$, where A_{bath} is the thermodynamic variable of the *bare bath*, with the interaction between the system and bath switched off. Hence it can be said that Seifert allocates the entire interaction energy to the system. These definitions of energy and entropy are further bootstrapped to nonequilibrium situations [5], and fluctuation theorems and the Clausius inequality are subsequently established. The resulting formulas (the right column of Table I) in strongly coupled regimes are markedly more complicated than those in weak coupling theory (the center column). These differences, however, disappear as the interaction Hamiltonian vanishes, and the HMF reduces to the bare system Hamiltonian.

Strasberg and Esposito [14] recently studied the strong coupling problem from the viewpoint of time-scale separation (TSS). They consider a system involving both slow and fast variables. By assuming fast variables in conditional equilibrium, they show that Seifert's theory can be derived by averaging out the fast variables. Furthermore, they proposed a definition of total entropy production in terms of the relative entropy, which is a variation of the entropy production defined in Ref. [27], and show that it is equivalent to the entropy production in Seifert's theory. The conditional equilibrium of the bath also allows one to prove the positivity of the instantaneous rate of total entropy production, rather than the positivity of the total entropy production of an entire process. The importance of TSS has long been known. It was invoked heuristically to justify adiabatic approximation [32,33], Markov modeling [34], and dimensional reduction of dynamic theories [35,36].

Jarzynski [7] developed a more comprehensive (and hence more complex) theory for strong coupling thermodynamics and systematically discussed the definitions of internal energy, entropy, volume, pressure, enthalpy, and Gibbs free energy. The formalism was established around the concept of volume, whose definition is somewhat arbitrary. All other thermodynamic variables are uniquely fixed by thermodynamic consistency once the system volume is (arbitrarily) defined. Jarzynski further showed that Seifert's theory is a special case of his (Jarzynski's) framework, i.e., the "partial molar representation." Jarzynski discussed in great detail the "bare representation," where the system enthalpy coincides with the HMF. The total entropy production is, however, the same in both representations. Jarzynski made an analogy between the arbitrariness in the definition of thermodynamic variables in the strong coupling regime and the gauge degree of freedom in electromagnetism, which was criticized by Talkner and Hänggi [8].

The main purpose of this paper is to show that, with TSS and the ensuing conditional equilibrium of bath variables, a much simpler thermodynamic theory can be developed for strongly coupled small classical systems. More specifically, we will show that by identifying the *Hamiltonian of mean force* (HMF) as the system Hamiltonian, and relegating the remaining part of the total Hamiltonian to the bath, both the equilibrium ensemble theory and the weak coupling theory of stochastic thermodynamics remain applicable, *almost verbatim*, in the strong coupling regime. Work and heat, entropy, and energy all retain the same definitions and the same physical meanings as in the weak coupling theory, as long as the bath entropy understood as conditioned on the system state. Fluctuation theorems, the Jarzynski equality, and the Clausius inequality can all be proved using nonlinear Langevin dynamics [37,38], whose validity relies on TSS but not on the strength of coupling. Using the conditional equilibrium nature of the bath, it can be rigorously demonstrated that $dS - \beta đQ$ is equal to the entropy change of the universe, which establishes the meaning of the Clausius inequality as increasing the total entropy. Finally, we will also show that our theory, though significantly simpler, is consistent with all previous theories, in the sense that the total entropy productions in all theories are mathematically equivalent. Summarizing, we achieve a natural unification of thermodynamics and stochastic thermodynamics at both weak and strong coupling regimes.

A logical consequence of TSS is that the dynamic evolution of slow variables can be modeled as a Markov process, such as Langevin dynamics with white noise. In the strongly coupled regime, the noises are, however, generically multiplicative. In a complementary paper [38], two of us develop a theory of stochastic thermodynamics using nonlinear Ito-Langevin dynamics, establish its covariance property, and derive the Crooks fluctuation theorem, Jarzynsk equality, and Clausius inequality. The definitions of thermodynamic quantities are identical in this paper and in Ref. [38], if we take $g_{ij} = \delta_{ij}$ in Ref. [38]. (The theory in Ref. [38] was developed for Langevin dynamics on an arbitrary Riemannian manifold with invariant volume measure $\sqrt{g}\, d^d x$, whereas in this paper, we consider Hamiltonian systems with Liouville measure $\prod_i dp_i dq_i$.) The combination of these two works provides a covariant theory of thermodynamics and stochastic thermodynamics for systems strongly interacting with a single heat bath, with TSS as the only assumption.

The remainder of this paper is organized as follows. In Sec. II, we introduce our decomposition of the total Hamiltonian and discuss the equilibrium thermodynamic properties of strongly coupled systems. In Sec. III, we discuss the nonequilibrium thermodynamic properties of the system. Work and heat retain the same definitions and same physical meanings as in the weak coupling theory, i.e., the energy changes of the combined system and of the bath, respectively. In Sec. IV, we discuss the connection between heat and entropy change of the bath, conditioned on the slow variables. In Sec. V, we compare our theory with previous theories by Seifert [5], by Talkner and Hänggi [6,8], by Jarzynski [7], and by Strasberg and Esposito [14] and show that they are all equivalent. We will also discuss a simple scenario where the present theory fits better with the common intuition about

system entropy and heat. In Sec. VI we make concluding remarks.

II. EQUILIBRIUM THEORY

In this section, we shall demonstrate that by identifying the HMF as the system Hamiltonian and the remainder of the total Hamiltonian as the bath Hamiltonian, canonical ensemble theory can be straightforwardly adapted to describe the equilibrium properties of systems that are strongly coupled to their baths. There is also a related decomposition of total thermodynamic quantities into system parts and bath parts. The bath free energy turns out to be the same as that of a bare bath and is independent of the state of slow variables or of the external control parameter.

A. Decomposition of the total Hamiltonian

We shall use $\mathbf{X}$, $\mathbf{Y}$ to denote fast and slow variables and $\boldsymbol{x}, \boldsymbol{y}$ to denote their values. We shall also call $\mathbf{X}$ the *system* and $\mathbf{Y}$ the *bath*. Let the total Hamiltonian be

$$H_{\mathbf{XY}}(\boldsymbol{x}, \boldsymbol{y}; \lambda) = H^0_{\mathbf{X}}(\boldsymbol{x}; \lambda) + H_{\mathbf{Y}}(\boldsymbol{y}) + H^0_I(\boldsymbol{x}, \boldsymbol{y}; \lambda), \quad (2.1)$$

where $H^0_{\mathbf{X}}(\boldsymbol{x}; \lambda)$ and $H_{\mathbf{Y}}(\boldsymbol{y})$ are the *bare system Hamiltonian* and *bare bath Hamiltonian*, whereas $H^0_I(\boldsymbol{x}, \boldsymbol{y}; \lambda)$ is the *bare interaction*. Note that every term on the right-hand side is independent of temperature, and the bare bath Hamiltonian $H_{\mathbf{Y}}(\boldsymbol{y})$ is independent of λ. Our starting point, Eq. (2.1), is more general than those in Ref. [5–7], where the bare interaction $H^0_I(\boldsymbol{x}, \boldsymbol{y}; \lambda)$ is assumed to be independent of λ.

Throughout this paper, we shall assume that $\mathbf{XY}$ is weakly interacting with a much larger superbath whose dynamics is even faster than $\mathbf{Y}$. We will call $\mathbf{YZ}$ the *environment* and $\mathbf{XYZ}$ the *universe*. We shall use $\int_{\boldsymbol{y}} \equiv \int d^N y$ to denote integration over $\boldsymbol{y}$ and similar notation for integration over $\boldsymbol{x}$ and $\mathbf{z}$. These notations are especially useful when we dealing with integrals of differential forms. Let $T = 1/\beta$ be the temperature, which is assumed to be fixed throughout this paper. We shall set the Boltzmann constant $k_B = 1$, and hence all entropies are dimensionless.

We shall define the *system Hamiltonian* $H_{\mathbf{X}}(\boldsymbol{x}; \lambda, \beta)$ and *interaction Hamiltonian* $H_I(\boldsymbol{x}, \boldsymbol{y}; \lambda, \beta)$ as

$$H_{\mathbf{X}}(\boldsymbol{x}; \lambda, \beta) = H^0_{\mathbf{X}}(\boldsymbol{x}) - T \ln \frac{\int_{\boldsymbol{y}} e^{-\beta(H_{\mathbf{Y}} + H^0_I)}}{\int_{\boldsymbol{y}} e^{-\beta H_{\mathbf{Y}}}}$$
$$= -T \ln \frac{\int_{\boldsymbol{y}} e^{-\beta H_{\mathbf{XY}}}}{\int_{\boldsymbol{y}} e^{-\beta H_{\mathbf{Y}}}}, \quad (2.2)$$

$$H_I(\boldsymbol{x}, \boldsymbol{y}; \lambda, \beta) = H^0_I(\boldsymbol{x}, \boldsymbol{y}; \lambda) + T \ln \frac{\int_{\boldsymbol{y}} e^{-\beta H_{\mathbf{Y}}}}{\int_{\boldsymbol{y}} e^{-\beta(H_{\mathbf{Y}} + H^0_I)}}, \quad (2.3)$$

both of which depend on β and λ. Note that $H_{\mathbf{X}}(\boldsymbol{x}; \lambda, \beta)$ is precisely the *Hamiltonian of mean force* (HMF) defined and used in previous works [5,6,16,30,39].

We now obtain a new decomposition of $H_{\mathbf{XY}}$:

$$H_{\mathbf{XY}}(\boldsymbol{x}, \boldsymbol{y}; \lambda) = H_{\mathbf{X}}(\boldsymbol{x}; \lambda, \beta) + H_{\mathbf{Y}}(\boldsymbol{y}) + H_I(\boldsymbol{x}, \boldsymbol{y}; \lambda, \beta). \quad (2.4a)$$

Note that even though both $H_{\mathbf{X}}$ and H_I depend on β, the total Hamiltonian on the left-hand side of Eq. (2.4a) is independent of β. We further define the *bath Hamiltonian* as

$$H_{\text{bath}}(\boldsymbol{y}; \boldsymbol{x}, \lambda, \beta) \equiv H_{\mathbf{Y}}(\boldsymbol{y}) + H_I(\boldsymbol{x}, \boldsymbol{y}; \lambda, \beta) \quad (2.4b)$$

and rewrite Eq. (2.4a) as

$$H_{\mathbf{XY}}(\boldsymbol{x}, \boldsymbol{y}; \lambda) = H_{\mathbf{X}}(\boldsymbol{x}; \lambda, \beta) + H_{\text{bath}}(\boldsymbol{y}; \boldsymbol{x}, \lambda, \beta). \quad (2.4c)$$

We also define the *bath partition function* as

$$Z_{\mathbf{Y}}(\boldsymbol{x}, \lambda, \beta) = \int_{\boldsymbol{y}} e^{-\beta H_{\text{bath}}(\boldsymbol{y}; \boldsymbol{x}, \lambda, \beta)}, \quad (2.5)$$

which is conditioned on $\mathbf{X} = \boldsymbol{x}$ and generally also depends on both $\boldsymbol{x}$ and λ. Using Eqs. (2.4b) and (2.3), we easily see that

$$Z_{\mathbf{Y}}(\boldsymbol{x}, \lambda, \beta) = \int_{\boldsymbol{y}} e^{-\beta(H_{\mathbf{Y}} + H^0_I)} \frac{\int_{\boldsymbol{y}} e^{-\beta H_{\mathbf{Y}}}}{\int_{\boldsymbol{y}} e^{-\beta(H_{\mathbf{Y}} + H^0_I)}}$$
$$= \int_{\boldsymbol{y}} e^{-\beta H_{\mathbf{Y}}} \equiv Z^0_{\mathbf{Y}}(\beta), \quad (2.6)$$

where $Z^0_{\mathbf{Y}}(\beta)$ is the partition function of the bare bath, with the interaction Hamiltonian between $\mathbf{X}$ and $\mathbf{Y}$ completely switched off.

Hence the bath partition function $Z_{\mathbf{Y}}(\boldsymbol{x}, \lambda, \beta)$ as defined by Eq. (2.5) is independent of $\boldsymbol{x}$ and λ:

$$\frac{\partial Z_{\mathbf{Y}}(\boldsymbol{x}, \lambda, \beta)}{\partial \boldsymbol{x}} = \frac{\partial Z_{\mathbf{Y}}(\boldsymbol{x}, \lambda, \beta)}{\partial \lambda} = 0, \quad (2.7)$$

and we shall from now on simply write it as $Z_{\mathbf{Y}}(\beta)$. Equation (2.7) will play a very significant role in our theory.

B. Conditional equilibrium of the bath

In an intermediate time scale, the fast variables equilibrate, whereas the slow variables barely change. Hence $\mathbf{Y}$ achieves equilibrium *conditioned on* $\mathbf{X} = \boldsymbol{x}$, described by the conditional Gibbs-Boltzmann distribution:

$$p^{\text{EQ}}_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{Z_{\mathbf{Y}}(\beta)} e^{-\beta H_{\text{bath}}(\boldsymbol{y}; \boldsymbol{x}, \lambda, \beta)}, \quad (2.8a)$$

with $Z_{\mathbf{Y}}(\beta)$ defined in Eq. (2.5). We further define the conditional free energy of the bath:

$$F_{\mathbf{Y}}(\beta) \equiv -T \ln Z_{\mathbf{Y}}(\beta) = -T \ln Z^0_{\mathbf{Y}}(\beta) = F^0_{\mathbf{Y}}(\beta), \quad (2.8b)$$

where $F^0_{\mathbf{Y}}(\beta)$ is the free energy of the bare bath. Equations (2.8a) and (2.8b) define a *conditional canonical ensemble*, which describes the equilibrium physics of the fast variables in the intermediate time scales, during which the slow variables change very little. In this ensemble theory, $\boldsymbol{x}$ serves as a parameter, just like λ and β.

The internal energy and entropy of the bath in the conditional equilibrium state are defined in a standard way:

$$E_{\mathbf{Y}}(\boldsymbol{x}) = \int_{\boldsymbol{y}} p^{\text{EQ}}_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{y}|\boldsymbol{x}) H_{\text{bath}}(\boldsymbol{y}; \boldsymbol{x}, \lambda, \beta), \quad (2.9a)$$

$$S_{\mathbf{Y}|\mathbf{X}=\boldsymbol{x}} = -\int_{\boldsymbol{y}} p^{\text{EQ}}_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{y}|\boldsymbol{x}) \ln p^{\text{EQ}}_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{y}|\boldsymbol{x}), \quad (2.9b)$$

which are related to the free energy $F_{\mathbf{Y}}(\beta)$ via

$$F_{\mathbf{Y}}(\beta) = E_{\mathbf{Y}}(\boldsymbol{x}) - T S_{\mathbf{Y}|\mathbf{X}=\boldsymbol{x}}. \quad (2.9c)$$

$S_{\mathbf{Y}|\mathbf{X}=\boldsymbol{x}}$ is known in information theory [40] as *the conditional Shannon entropy of* $\mathbf{Y}$ *given* $\mathbf{X}=\boldsymbol{x}$. Note that even though $F_{\mathbf{Y}}(\beta)$ does not depend on $\boldsymbol{x}$ and λ, both $E_{\mathbf{Y}}(\boldsymbol{x})$ and $S_{\mathbf{Y}|\mathbf{X}=\boldsymbol{x}}$ do.

Even though the free energy of the bath conditioned on the system state is equal to that of the bare bath, there are important differences between other thermodynamic quantities of the bath and the bare bath. For example, the internal energy and entropy of the bare bath are given by

$$E^0_{\mathbf{Y}} = \int_{\boldsymbol{y}} \frac{e^{-\beta H_{\mathbf{Y}}}}{Z^0_{\mathbf{Y}}(\beta)} H_{\mathbf{Y}}(\boldsymbol{y}), \tag{2.10a}$$

$$S^0_{\mathbf{Y}} = -\int_{\boldsymbol{y}} \frac{e^{-\beta H_{\mathbf{Y}}}}{Z^0_{\mathbf{Y}}(\beta)} \ln \frac{e^{-\beta H_{\mathbf{Y}}}}{Z^0_{\mathbf{Y}}(\beta)}, \tag{2.10b}$$

respectively, which are manifestly different from Eqs. (2.9a) and (2.9b).

C. Joint equilibrium of the system and bath

In long time scales, $\mathbf{XY}$ achieve a joint equilibrium, which is described by the joint Gibbs-Boltzmann distribution

$$p^{\mathrm{EQ}}_{\mathbf{XY}}(\boldsymbol{x},\boldsymbol{y}) = \frac{e^{-\beta H_{\mathbf{XY}}(\boldsymbol{x},\boldsymbol{y};\lambda)}}{Z_{\mathbf{XY}}(\beta,\lambda)}, \tag{2.11}$$

where $Z_{\mathbf{XY}}(\beta,\lambda)$ is the canonical joint partition function

$$\begin{aligned} Z_{\mathbf{XY}}(\beta,\lambda) &= \int_{\boldsymbol{xy}} e^{-\beta H_{\mathbf{XY}}} \\ &= \int_{\boldsymbol{xy}} e^{-\beta H_{\mathbf{X}} - \beta H_{\mathrm{bath}}}. \end{aligned} \tag{2.12a}$$

From this we can obtain various thermodynamic quantities for this *joint canonical ensemble* in a standard way:

$$F_{\mathbf{XY}}(\beta,\lambda) = -T \ln Z_{\mathbf{XY}}(\beta,\lambda), \tag{2.12b}$$

$$E_{\mathbf{XY}}(\beta,\lambda) = \int_{\boldsymbol{x},\boldsymbol{y}} p^{\mathrm{EQ}}_{\mathbf{XY}}(\boldsymbol{x},\boldsymbol{y}) H_{\mathbf{XY}}(\boldsymbol{x},\boldsymbol{y};\lambda), \tag{2.12c}$$

$$S_{\mathbf{XY}}(\beta,\lambda) = -\int_{\boldsymbol{x},\boldsymbol{y}} p^{\mathrm{EQ}}_{\mathbf{XY}}(\boldsymbol{x},\boldsymbol{y}) \ln p^{\mathrm{EQ}}_{\mathbf{XY}}(\boldsymbol{x},\boldsymbol{y}), \tag{2.12d}$$

$$F_{\mathbf{XY}}(\beta,\lambda) = E_{\mathbf{XY}} - T S_{\mathbf{XY}}. \tag{2.12e}$$

The joint canonical ensemble describes the equilibrium statistical properties of both slow and fast variables.

D. Reduced equilibrium of the system

We may also study the equilibrium distribution of slow variables alone. This *reduced canonical distribution* can be obtained from Eq. (2.11) by integrating out the fast variables:

$$\begin{aligned} p^{\mathrm{EQ}}_{\mathbf{X}}(\boldsymbol{x}) &= \int_{\boldsymbol{y}} p^{\mathrm{EQ}}_{\mathbf{XY}}(\boldsymbol{x},\boldsymbol{y}) \\ &= \frac{1}{Z_{\mathbf{XY}}(\beta,\lambda)} \int_{\boldsymbol{y}} e^{-\beta H_{\mathbf{X}} - \beta H_{\mathrm{bath}}} \\ &= \frac{Z_{\mathbf{Y}}(\beta)}{Z_{\mathbf{XY}}(\beta,\lambda)} e^{-\beta H_{\mathbf{X}}}, \end{aligned} \tag{2.13}$$

where we used Eq. (2.5) and the fact that $Z_{\mathbf{Y}}(\beta)$ is independent of $\boldsymbol{x}$. Hence the equilibrium distribution of $\mathbf{X}$ is canonical with respect to the system Hamiltonian $H_{\mathbf{X}}(\boldsymbol{x};\lambda)$. This is, of course, well known, since we have defined $H_{\mathbf{X}}(\boldsymbol{x};\lambda)$ as the HMF.

It is then convenient to define the partition function of slow variables,

$$Z_{\mathbf{X}}(\beta,\lambda) \equiv \int_{\boldsymbol{x}} e^{-\beta H_{\mathbf{X}}(\boldsymbol{x};\lambda,\beta)}, \tag{2.14}$$

so that Eq. (2.13) assumes the standard canonical form

$$p^{\mathrm{EQ}}_{\mathbf{X}}(\boldsymbol{x}) = \frac{1}{Z_{\mathbf{X}}(\beta,\lambda)} e^{-\beta H_{\mathbf{X}}}. \tag{2.15}$$

Integration of Eq. (2.13) then yields

$$Z_{\mathbf{XY}}(\beta,\lambda) = Z_{\mathbf{X}}(\beta,\lambda) Z_{\mathbf{Y}}(\beta). \tag{2.16}$$

The above results prompt us to construct a *reduced canonical ensemble* theory for the system, with free energy, internal energy, and entropy given by

$$F_{\mathbf{X}} = -T \ln Z_{\mathbf{X}}(\beta,\lambda), \tag{2.17a}$$

$$E_{\mathbf{X}} = \int_{\boldsymbol{x}} p^{\mathrm{EQ}}_{\mathbf{X}}(\boldsymbol{x}) H_{\mathbf{X}}(\boldsymbol{x};\lambda), \tag{2.17b}$$

$$S_{\mathbf{X}} = -\int_{\boldsymbol{x}} p^{\mathrm{EQ}}_{\mathbf{X}}(\boldsymbol{x}) \ln p^{\mathrm{EQ}}_{\mathbf{X}}(\boldsymbol{x}), \tag{2.17c}$$

$$F_{\mathbf{X}} = E_{\mathbf{X}} - T S_{\mathbf{X}}. \tag{2.17d}$$

These definitions of system energy and entropy are manifestly different from the strong coupling theory in Refs. [5,6,8], even though the free energy is the same in the two theories.

E. Decomposition of thermodynamic variables

Comparing Eqs. (2.17a)–(2.17d) with Eqs. (2.9a)–(2.9c) and (2.12a)–(2.12e), we find the following decomposition of total thermodynamic quantities into system parts and bath parts:

$$F_{\mathbf{XY}}(\beta,\lambda) = F_{\mathbf{X}}(\beta,\lambda) + F_{\mathbf{Y}}(\beta), \tag{2.18a}$$

$$E_{\mathbf{XY}} = E_{\mathbf{X}} + \langle E_{\mathbf{Y}}(\boldsymbol{x}) \rangle_{\mathbf{X}}, \tag{2.18b}$$

$$S_{\mathbf{XY}} = S_{\mathbf{X}} + S_{\mathbf{Y}|\mathbf{X}}, \tag{2.18c}$$

where $\langle E_{\mathbf{Y}}(\boldsymbol{x}) \rangle_{\mathbf{X}}$ and $S_{\mathbf{Y}|\mathbf{X}}$ are averages of $E_{\mathbf{Y}}(\boldsymbol{x})$ and $S_{\mathbf{Y}|\mathbf{X}=\boldsymbol{x}}$, respectively, over fluctuations of $\mathbf{X}$,

$$\langle E_{\mathbf{Y}}(\boldsymbol{x}) \rangle_{\mathbf{X}} = \int_{\boldsymbol{x}} p^{\mathrm{EQ}}_{\mathbf{X}}(\boldsymbol{x}) E_{\mathbf{Y}}(\boldsymbol{x}), \tag{2.19a}$$

$$S_{\mathbf{Y}|\mathbf{X}} = \langle S_{\mathbf{Y}|\mathbf{X}=\boldsymbol{x}} \rangle_{\mathbf{X}} = \int_{\boldsymbol{x}} p^{\mathrm{EQ}}_{\mathbf{X}}(\boldsymbol{x}) S_{\mathbf{Y}|\mathbf{X}=\boldsymbol{x}}. \tag{2.19b}$$

$S_{\mathbf{Y}|\mathbf{X}}$ is called the *conditional Shannon entropy of* $\mathbf{Y}$ *given* $\mathbf{X}$ in information theory [40]. Note the subtle differences between the name for $S_{\mathbf{Y}|\mathbf{X}}$ and the name for $S_{\mathbf{Y}|\mathbf{X}=\boldsymbol{x}}$.

There are numerous pleasant features of the equilibrium thermodynamic theory developed here: Firstly, all equilibrium distributions are Gibbs-Boltzmann distributions with respect to the corresponding Hamiltonian. Secondly, all entropies are Gibbs-Shannon entropies with respect to the corresponding probability density function (pdfs). As a consequence, the formulas in Eqs. (2.8a) and (2.8b), (2.9a)–(2.9c), (2.12a)–(2.12e), and (2.17a)–(2.17d) are all the same as those in

canonical ensemble theory. These features are remarkable, since they indicate that standard canonical ensemble theory is applicable both to the system and to the bath, regardless of the strong interaction and correlation between them. Thirdly, Eq. (2.8b) says that the bath free energy $F_{\mathbf{Y}}(\beta)$ is equal to the bare bath free energy $F^0_{\mathbf{Y}}(\beta)$ and is independent of λ and $\boldsymbol{x}$. This feature leads to substantial conceptual simplification since we are only interested in the physics of slow variables. Consider, for example, that we immerse a DNA chain into an aqueous solvent, or stretch it in the solvent, or tune the interaction between a nanoengine and its environment. There is no need to worry about the change in the bath free energy, because it stays constant by construction.

All these convenient features follow from the particular decomposition of the total Hamiltonian equations (2.4a)–(2.4c). There are, however, some subtleties resulting from the temperature dependence of $H_{\mathbf{X}}$, which will be discussed in Sec. V. We shall also give a detailed comparison between our theory and the previous theories by Seifert [5], Talkner and Hänggi [6,8], and Jarzynski [7] in Sec. V.

III. NONEQUILIBRIUM THEORY

In this section, we shall show that with the HMF $H_{\mathbf{X}}$ identified as the fluctuating internal energy, the weak coupling theory of stochastic thermodynamics becomes applicable in the strong coupling regime.

A. Definitions of energy and entropy

The mission of stochastic thermodynamics starts with definitions of system thermodynamic variables in general nonequilibrium situations. We define the fluctuating internal energy of the system as $H_{\mathbf{X}}(\boldsymbol{x};\lambda,\beta)$, the HMF. The nonequilibrium internal energy is then defined as the ensemble average of $H_{\mathbf{X}}$:

$$E_{\mathbf{X}}[p_{\mathbf{X}}] \equiv -\int_{\boldsymbol{x}} p_{\mathbf{X}} H_{\mathbf{X}}. \tag{3.1}$$

Throughout this paper we use $A[p_{\mathbf{X}}]$ to denote a nonequilibrium thermodynamic variable, to distinguish it from the equilibrium version A. We also define the system entropy as the Gibbs-Shannon entropy:

$$S_{\mathbf{X}}[p_{\mathbf{X}}] \equiv -\int_{\boldsymbol{x}} p_{\mathbf{X}}(\boldsymbol{x},t)\ln p_{\mathbf{X}}(\boldsymbol{x},t). \tag{3.2}$$

We shall not need to define stochastic entropy in this paper. The nonequilibrium free energy of the system is also defined in the standard way:

$$\begin{aligned} F_{\mathbf{X}}[p_{\mathbf{X}}] &= E_{\mathbf{X}}[p_{\mathbf{X}}] - T\,S_{\mathbf{X}}[p_{\mathbf{X}}] \\ &= \int_{\boldsymbol{x}} p_{\mathbf{X}}(H_{\mathbf{X}} + T\,\ln p_{\mathbf{X}}), \end{aligned} \tag{3.3}$$

which turns out to be the same as the free energy defined in several previous theories [5,6,8,14].

Note that these definitions of nonequilibrium entropy, energy, and free energy are identical to those in weak coupling theory, with $H_{\mathbf{X}}$ understood as the system Hamiltonian. For equilibrium state $p_{\mathbf{X}} = p^{\mathrm{EQ}}_{\mathbf{X}}$, these thermodynamic variables reduce to their equilibrium counterparts, Eqs. (2.17b), (2.17c), and (2.17a), respectively.

B. Work and heat at the trajectory level

Let us now discuss differential work and heat at the trajectory level of system variables.

The Hamiltonian of the *universe*, including system, bath, and superbath, is given by

$$\begin{aligned} H_{\mathbf{XYZ}} &= H_{\mathbf{XY}} + H_{\mathbf{Z}} \\ &= H_{\mathbf{X}} + H_{\mathrm{bath}} + H_{\mathbf{Z}}, \end{aligned} \tag{3.4}$$

with $H_{\mathbf{XY}}$ given by Eqs. (2.4a)–(2.4c). We assume that the interaction between $\mathbf{XY}$ and $\mathbf{Z}$ is negligibly small but nonetheless is strong enough to drive thermal equilibration between $\mathbf{XY}$ and $\mathbf{Z}$.

We consider a microscopic process with infinitesimal duration dt, where $\boldsymbol{x},\boldsymbol{y},\mathbf{z}$ and λ change by $d\boldsymbol{x},d\boldsymbol{y},d\mathbf{z}$ and $d\lambda$. Whereas $d\lambda$ is externally controlled, $d\boldsymbol{x},d\boldsymbol{y},d\mathbf{z}$ are determined by evolution of the Hamiltonian dynamics. As is generally adopted in stochastic thermodynamics, work is defined as the change in the total energy of the universe:

$$đ\tilde{W} = dH_{\mathbf{XYZ}}. \tag{3.5}$$

For now we shall assume that $\boldsymbol{x},\boldsymbol{y},\mathbf{z},\lambda$ are all smooth functions of t [41]. We can then expand Eq. (3.5) in terms of $d\boldsymbol{x},\ldots,d\lambda$ up to the first order. The coefficients are just the partial derivatives of $H_{\mathbf{XYZ}}$ with respect to $d\boldsymbol{x},\ldots,d\lambda$. Now note that the universe $\mathbf{XYZ}$ is thermally closed. Hence, if λ is fixed, $H_{\mathbf{XYZ}}$ must be conserved. In other words, Eq. (3.5) can change only due to λ:

$$đ\tilde{W} = \frac{\partial H_{\mathbf{XYZ}}}{\partial\lambda}d\lambda \equiv d_\lambda H_{\mathbf{XYZ}}. \tag{3.6}$$

Further using Eqs. (3.4) and (2.4a), we can rewrite the preceding equation as

$$đ\tilde{W} = d_\lambda(H_{\mathbf{XY}} + H_{\mathbf{Z}}) = d_\lambda H_{\mathbf{X}} + d_\lambda H_I, \tag{3.7}$$

where in the last equality we have used the fact that both $H_{\mathbf{Y}}$ and $H_{\mathbf{Z}}$ are independent of λ. Hence the microscopic work $đ\tilde{W}$ is independent of the state of the superbath.

Note that the work $đ\tilde{W}$ as given by Eq. (3.7) depends on $\boldsymbol{x},\boldsymbol{y},\lambda,d\lambda$. In stochastic thermodynamics, we keep track of the dynamic evolution of $\boldsymbol{x}$ but not of $\boldsymbol{y}$. Hence, to obtain the differential work at the trajectory level of system variables, we need to average Eq. (3.7) over the conditional equilibrium as given by Eq. (2.8a):

$$\begin{aligned} đW &= \int_{\boldsymbol{y}} p^{\mathrm{EQ}}_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{y}|\boldsymbol{x})\,đ\tilde{W} \\ &= \int_{\boldsymbol{y}} p^{\mathrm{EQ}}_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{y}|\boldsymbol{x})(d_\lambda H_{\mathbf{X}} + d_\lambda H_I). \end{aligned} \tag{3.8}$$

This equation and many analogous equations below are understood as the volume integral of differential forms. Be careful not to confuse the differential forms $đ\tilde{W}$, $d_\lambda H_{\mathbf{X}}$, etc., with the volume measure $d^N y$ which is hidden in $\int_{\boldsymbol{y}}$.

Now, taking the λ differential of Eq. (2.5), and further using Eq. (2.7), we find

$$\int_y p^{\rm EQ}_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{y}|\boldsymbol{x}) d_\lambda H_I = 0. \tag{3.9}$$

Hence Eq. (3.8) reduces to

$$đ\mathscr{W} = d_\lambda H_{\mathbf{X}} = \frac{\partial H_{\mathbf{X}}}{\partial \lambda} d\lambda. \tag{3.10}$$

Hence, even though the interaction Hamiltonian H_I may be tuned externally, the work $đ\mathscr{W}$ at the trajectory level is nonetheless independent of H_I.

Taking the differential of Eq. (3.4) and using Eq. (3.5), we obtain

$$dH_{\mathbf{X}} = đ\tilde{\mathscr{W}} - d(H_B + H_{\mathbf{Z}}). \tag{3.11}$$

As in the above, we take the average Eq. (3.11) over fluctuations of $\mathbf{YZ}$, which results in

$$dH_{\mathbf{X}} = đ\mathscr{W} + đ\mathcal{Q}, \tag{3.12}$$

$$đ\mathcal{Q} \equiv -d\langle H_B + H_{\mathbf{Z}} \rangle_{\mathbf{YZ}}, \tag{3.13}$$

where $\langle \cdot \rangle_{\mathbf{YZ}}$ means the average over $\mathbf{YZ}$ and $đ\mathcal{Q}$ is the differential heat at the trajectory level of the system variables. Since $H_{\mathbf{X}}$ is defined as the fluctuating internal energy and $đ\mathscr{W}$ is the work at the trajectory level, Eq. (3.12) can be interpreted as the first law at the trajectory level if $đ\mathcal{Q} = -d(H_B + H_{\mathbf{Z}})$ is interpreted as the *heat at the trajectory level*. Equation (3.13) then says that the heat $đ\mathcal{Q}$ is negative the average energy variation of the environment $\mathbf{YZ}$. Such an interpretation of heat is exactly the same as that in the weak coupling stochastic thermodynamics.

However, the differential of fluctuating internal energy $dH_{\mathbf{X}}$ can be written as the sum of $d_\lambda H$ and $d_{\boldsymbol{x}} H_{\mathbf{X}}$:

$$dH_{\mathbf{X}} = d_\lambda H_{\mathbf{X}} + d_{\boldsymbol{x}} H_{\mathbf{X}}. \tag{3.14}$$

Comparing this with Eq. (3.12), we obtain an alternative expression for $đ\mathcal{Q}$:

$$đ\mathcal{Q} \equiv dH_{\mathbf{X}} - đ\mathscr{W} = d_{\boldsymbol{x}} H_{\mathbf{X}}, \tag{3.15}$$

which must be equivalent to Eq. (3.13). It is tempting to rewrite $d_{\boldsymbol{x}} H_{\mathbf{X}}$ in terms of partial derivatives

$$d_{\boldsymbol{x}} H_{\mathbf{X}} = \frac{\partial H_{\mathbf{X}}}{\partial \boldsymbol{x}} d\boldsymbol{x}. \tag{3.16}$$

This is, however, valid only if $\boldsymbol{x}(t)$ is differentiable so that $d\boldsymbol{x}$ is linear in dt. In the limit of time-scale separation, we expect that a typical path of slow variables $\boldsymbol{x}(t)$ becomes that of Brownian motion, which is everywhere continuous but nondifferentiable. As a consequence, $d\boldsymbol{x}(t)$ scales as $\sqrt{dt}$ (Ito's formula), and we need to expand $d_{\boldsymbol{x}} H_{\mathbf{X}}$ up to the second order in $d\boldsymbol{x}$, if the product on the right-hand side of Eq. (3.16) is defined in Ito's sense. We can also interpret the product on the right-hand side of Eq. (3.16) in Stratonovich's sense. Then Eq. (3.16) remains valid even if $\boldsymbol{x}(t)$ is a typical path of Brownian motion. In this paper, we shall not write $d_{\boldsymbol{x}} H_{\mathbf{X}}$ in terms of partial derivatives, so that we do not need to worry about the issue of stochastic calculus.

Note that the definitions of work and heat at the trajectory level, Eqs. (3.10) and (3.15), are the same as those in the weak coupling theory.

C. Work and heat at the ensemble level

To obtain work and heat at the ensemble level, we need to average corresponding objects at the trajectory level over the (generally out-of-equilibrium) statistical distribution of dynamic trajectories of $\mathbf{X}$. This is a rather nontrivial task. Luckily, $đ\mathscr{W}$ as given by Eq. (3.10) is independent of $d\boldsymbol{x}$. Hence we do not need to know the pdf of $d\boldsymbol{x}$, but only need to average Eq. (3.10) over statistical distribution $p_{\mathbf{X}}(\boldsymbol{x}, t)$ at time t, and obtain the differential work $đW$ at the ensemble level:

$$đW = \int_{\boldsymbol{x}} p_{\mathbf{X}} d_\lambda H_{\mathbf{X}}. \tag{3.17}$$

Now we want to take the ensemble average of heat, Eq. (3.15), which does depend on $d\boldsymbol{x} \equiv \boldsymbol{x}(t + dt) - \boldsymbol{x}(t)$, whose distribution is not encoded in the instantaneous distribution $p_{\mathbf{X}}(\boldsymbol{x}, t)$. A dynamic theory for $d\boldsymbol{x}$, such as nonlinear Langevin dynamics, would supply the necessary information. This route was pursued in the complementary work, Ref. [38]. Here, we take a detour by studying the average of $dH_{\mathbf{X}}$. Let $p_{\mathbf{X}}(\boldsymbol{x}, t)$ and $p_{\mathbf{X}}(\boldsymbol{x}, t + dt)$ be the pdfs of $\boldsymbol{x}$ at t and at $t + dt$, respectively, and $dp_{\mathbf{X}}(\boldsymbol{x}, t)$ be the differential of $p_{\mathbf{X}}(\boldsymbol{x}, t)$ as given by

$$\begin{aligned} dp_{\mathbf{X}}(\boldsymbol{x}, t) &= p_{\mathbf{X}}(\boldsymbol{x}, t + dt) - p_{\mathbf{X}}(\boldsymbol{x}, t) \\ &= \frac{\partial p_{\mathbf{X}}(\boldsymbol{x}, t)}{\partial t} dt. \end{aligned} \tag{3.18}$$

Let us calculate the differential of the internal energy as given by Eq. (3.1):

$$dE_{\mathbf{X}}[p_{\mathbf{X}}] = d\langle H_{\mathbf{X}} \rangle = d \int_{\boldsymbol{x}} H_{\mathbf{X}} p_{\mathbf{X}}. \tag{3.19}$$

Since $\boldsymbol{x}$ is integrated over on the right-hand side, the differential d is due to changes in λ and in $p(\boldsymbol{x}, t)$:

$$d\langle H_{\mathbf{X}} \rangle = \int_{\boldsymbol{x}} (d_\lambda H_{\mathbf{X}})\, p_{\mathbf{X}} + \int_{\boldsymbol{x}} H_{\mathbf{X}}\, dp_{\mathbf{X}}. \tag{3.20}$$

However, the first term on the right-hand side is just the work at the ensemble level, as we defined in Eq. (3.17). Hence the second term must be the heat at the ensemble level,

$$đQ = \int_{\boldsymbol{x}} H_{\mathbf{X}} dp_{\mathbf{X}} = \langle đ\mathcal{Q} \rangle, \tag{3.21}$$

and Eq. (3.20) becomes the first law at the ensemble level:

$$dE_{\mathbf{X}}[p_{\mathbf{X}}] = đW + đQ. \tag{3.22}$$

The definitions of work and heat at the ensemble level, Eqs. (3.17) and (3.21), are again the same as those in the weak coupling theory of stochastic thermodynamics.

IV. PHYSICAL MEANINGS OF HEAT

In this section, we shall establish the connections between heat (both at the trajectory level and at the ensemble level) and entropy change of the environment, conditioned on the system state and possibly other thermodynamic variables. We shall

also discuss the physical meanings of the Clausius inequality and total entropy production. The results are again the same as those in the weak coupling theory, with the conditioning of slow variables properly taken into account.

A. Heat at the trajectory level

The universe **XYZ** is thermally closed and evolves according to Hamiltonian dynamics with the Hamiltonian given in Eq. (3.4). Due to TSS, with $\boldsymbol{x}$ fixed, the *environment* **YZ** is described by a microcanonical ensemble with fixed energy. We define the Boltzmann entropy of the environment as a function of its energy $E_{\mathbf{YZ}}$:

$$\begin{aligned} S_{\mathbf{YZ}}(E_{\mathbf{YZ}}) &\equiv \ln\ \Omega_{\mathbf{YZ}}(E_{\mathbf{YZ}}) \\ &\equiv \ln \int_{\boldsymbol{y},\mathbf{z}} \delta(H_{\text{bath}} + H_{\mathbf{Z}} - E_{\mathbf{YZ}}), \end{aligned} \tag{4.1}$$

where H_{bath} is defined in Eq. (2.4b), and $\Omega_{\mathbf{YZ}}(E_{\mathbf{YZ}})$ is the area of the **YZ** hypersurface with constant bath energy $E_{\mathbf{YZ}}$. Note that $S_{\mathbf{YZ}}(E_{\mathbf{YZ}})$ generally also depends on $\boldsymbol{x}, \lambda, \beta$ parametrically through H_{bath}. We shall, however, not explicitly display the parameters $\boldsymbol{x}, \lambda, \beta$, in order not to make the notations too cluttered. Strictly speaking, $S_{\mathbf{YZ}}(E_{\mathbf{YZ}})$ is *the Boltzmann entropy of the environment conditioned on* $\mathbf{X} = \boldsymbol{x}$.

Suppose that in the initial state the system is at $\boldsymbol{x}$ with external parameter λ, and the universe **XYZ** has total energy $E_{\mathbf{XYZ}}$. The energy of the environment is then $E_{\mathbf{YZ}} = E_{\mathbf{XYZ}} - H_{\mathbf{X}}$. In the final state the system is at $\boldsymbol{x} + d\boldsymbol{x}$ with external parameter $\lambda + d\lambda$, and the universe has total energy $E_{\mathbf{XYZ}} + đ\mathscr{W}$, with $đ\mathscr{W}$ given by Eq. (3.10). (Recall that the work is defined as the change in total energy.) The energy of the environment in the final state is then $E'_{\mathbf{YZ}} = E_{\mathbf{XYZ}} + đ\mathscr{W} - H_{\mathbf{X}} - dH_{\mathbf{X}}$, where $dH_{\mathbf{X}}$ is given by Eq. (3.14). The Boltzmann entropies of the environment in the initial and final states are hence

$$S_{\mathbf{YZ}}(E_{\mathbf{YZ}}) = S_{\mathbf{YZ}}(E_{\mathbf{XYZ}} - H_{\mathbf{X}}), \tag{4.2a}$$

$$S_{\mathbf{YZ}}(E'_{\mathbf{YZ}}) = S_{\mathbf{YZ}}(E_{\mathbf{XYZ}} + đ\mathscr{W} - H_{\mathbf{X}} - dH_{\mathbf{X}}), \tag{4.2b}$$

respectively. Note that $E_{\mathbf{XYZ}}$ is much larger than $dH_{\mathbf{X}}, đ\mathscr{W}$, because the size of the superbath is much larger than **XY**. Expanding Eq. (4.2b) in terms of $đ\mathscr{W}$ and $dH_{\mathbf{X}}$ and subtracting from it Eq. (4.2a), we obtain

$$\begin{aligned} dS_{\mathbf{YZ}}(E_{\mathbf{YZ}}) &= S_{\mathbf{YZ}}(E'_{\mathbf{YZ}}) - S_{\mathbf{YZ}}(E_{\mathbf{YZ}}) \\ &= \beta(đ\mathscr{W} - dH_{\mathbf{X}}) = -\beta đ\mathcal{Q}, \end{aligned} \tag{4.3}$$

where $\beta = \partial S_{\mathbf{YZ}}/\partial E_{\mathbf{YZ}}$ is the inverse temperature. Further using Eq. (3.15), we find

$$-\beta đ\mathcal{Q} = dS_{\mathbf{YZ}}(E_{\mathbf{YZ}}), \tag{4.4}$$

which establishes the connection between the differential heat $đ\mathcal{Q}$ at the level of the system trajectory and the differential of the environment Boltzmann entropy $dS_{\mathbf{YZ}}(E_{\mathbf{YZ}})$ conditioned on $\mathbf{X} = \boldsymbol{x}$.

B. Heat at the ensemble level and total entropy production

Recall that **XY** is in contact with a much larger superbath **Z** and that **Y** is always in conditional equilibrium. If the system is in a nonequilibrium state $p_{\mathbf{X}}(\boldsymbol{x})$, the joint pdf of **XY** is given by

$$p_{\mathbf{XY}}(\boldsymbol{x}, \boldsymbol{y}) = p_{\mathbf{X}}(\boldsymbol{x})\, p^{\text{EQ}}_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{y}|\boldsymbol{x}), \tag{4.5}$$

where $p^{\text{EQ}}_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{y}|\boldsymbol{x})$ is given in Eq. (2.8a). The nonequilibrium free energy for the system is already defined in Eq. (3.3). Let us similarly define the nonequilibrium free energy of the combined system **XY**:

$$F_{\mathbf{XY}}[p_{\mathbf{XY}}] \equiv \int_{\boldsymbol{x},\boldsymbol{y}} p_{\mathbf{XY}}(H_{\mathbf{XY}} + T\ \ln\ p_{\mathbf{XY}}). \tag{4.6}$$

For **XY**, there is no difference between the Hamiltonian and the Hamiltonian of mean force, since **XY** is in weak interaction with **Z**. Substituting Eq. (4.5) into Eq. (4.6), and using Eqs. (2.4b) and (2.8a) and (2.8b), we obtain

$$F_{\mathbf{XY}}[p_{\mathbf{XY}}] = F_{\mathbf{X}}[p_{\mathbf{X}}] + F_{\mathbf{Y}}(\beta), \tag{4.7}$$

which says that $F_{\mathbf{XY}}[p_{\mathbf{XY}}]$ and $F_{\mathbf{X}}[p_{\mathbf{X}}]$ differ only by an additive constant $F_{\mathbf{Y}}(\beta)$, which is, according to Eq. (2.7), independent of λ and $\boldsymbol{x}$ and hence needs to be worried about when we study nonequilibrium processes. Equation (4.7) is a nonequilibrium generalization of Eq. (2.18a).

Let us now consider variations of λ and $p_{\mathbf{X}}$ and study the resulting variation of free energies. Taking the differential of Eq. (3.3), we obtain

$$dF_{\mathbf{X}}[p_{\mathbf{X}}] = đW + đQ - T dS_{\mathbf{X}}[p_{\mathbf{X}}], \tag{4.8}$$

where $đW$ and $đQ$ are work and heat at the ensemble level, given in Eqs. (3.17) and (3.21), respectively. We can rewrite this result into

$$dS_{\mathbf{X}}[p_{\mathbf{X}}] - \beta đQ = \beta(đW - dF_{\mathbf{X}}[p_{\mathbf{X}}]). \tag{4.9}$$

We can also do a similar thing for $dF_{\mathbf{XY}}[p_{\mathbf{X}}]$, and we obtain an analogous result:

$$dS_{\mathbf{XY}}[p_{\mathbf{XY}}] - \beta đQ_{\mathbf{XY}} = \beta(đW_{\mathbf{XY}} - dF_{\mathbf{XY}}[p_{\mathbf{XY}}]), \tag{4.10}$$

where $đW_{\mathbf{XY}}$ and $đQ_{\mathbf{XY}}$ are the work and heat at the ensemble level of **XY**,

$$đW_{\mathbf{XY}} = \int_{\boldsymbol{x},\boldsymbol{y}} p_{\mathbf{XY}} d_\lambda H_{\mathbf{XY}}, \tag{4.11}$$

$$đQ_{\mathbf{XY}} = \int_{\boldsymbol{x},\boldsymbol{y}} d p_{\mathbf{XY}}\, H_{\mathbf{XY}}. \tag{4.12}$$

Using Eqs. (4.5) and (3.9) in Eq. (4.11), we see that

$$đW_{\mathbf{XY}} = đW. \tag{4.13}$$

Taking the differential of Eq. (4.7), we find

$$dF_{\mathbf{X}}[p_{\mathbf{X}}] = dF_{\mathbf{XY}}[p_{\mathbf{XY}}]. \tag{4.14}$$

Combining the preceding two equations with Eqs. (4.9) and (4.10), we find

$$\begin{aligned} dS_{\mathbf{XY}}[p_{\mathbf{XY}}] - \beta đQ_{\mathbf{XY}} &= \beta(đW_{\mathbf{XY}} - dF_{\mathbf{XY}}[p_{\mathbf{XY}}]) \\ &= dS_{\mathbf{X}}[p_{\mathbf{X}}] - \beta đQ \\ &= \beta(đW - dF_{\mathbf{X}}[p_{\mathbf{X}}]). \end{aligned} \tag{4.15}$$

Now recall that **XY** is weakly coupled to the superbath **Z** and hence the weak coupling theory of stochastic thermodynamics is applicable. It tells us that $dS_{\mathbf{XY}}[p_{\mathbf{XY}}] - \beta đQ_{\mathbf{XY}}$ is positive definite and can be interpreted as the change in

TABLE I. Major formulas of thermodynamics and stochastic thermodynamics. According to the present theory, formulas in the center column are applicable both in the weak coupling regime and in the strong coupling regime, with $H_{\mathbf{X}}$ being the Hamiltonian of mean force. The formulas in Seifert's theory [5], which are substantially more complex, are shown in the right column. The differences disappear in the weak coupling limit, where $H_{\mathbf{X}}$ reduces to the bare system Hamiltonian, which is independent of β. FT, function theorem; ineq., inequality; TM, thermodynamic.

TM quantities and laws	Weak coupling theory and the present theory	Seifert's strong coupling theory
Fluctuating internal energy	$H_{\mathbf{X}}$	$\tilde{H}_{\mathbf{X}} \equiv \partial_\beta \beta H_{\mathbf{X}}$
Internal energy	$E_{\mathbf{X}}[p_{\mathbf{X}}] = \int_{\boldsymbol{x}} p_{\mathbf{X}} H_{\mathbf{X}}$	$\tilde{E}_{\mathbf{X}}[p_{\mathbf{X}}] = \int_{\boldsymbol{x}} p_{\mathbf{X}} \partial_\beta \beta H_{\mathbf{X}}$
Entropy	$S_{\mathbf{X}}[p_{\mathbf{X}}] = -\int_{\boldsymbol{x}} p_{\mathbf{X}} \ln p_{\mathbf{X}}$	$\tilde{S}_{\mathbf{X}}[p_{\mathbf{X}}] = \int_{\boldsymbol{x}} p_{\mathbf{X}}(-\ln p_{\mathbf{X}} + \beta^2 \partial_\beta H_{\mathbf{X}})$
Free energy	$F_{\mathbf{X}} = E_{\mathbf{X}} - T S_{\mathbf{X}}$	$F_{\mathbf{X}} = \tilde{E}_{\mathbf{X}} - T \tilde{S}_{\mathbf{X}} = E_{\mathbf{X}} - T S_{\mathbf{X}}$
Work at trajectory level	$đ\mathscr{W} = d_\lambda H_{\mathbf{X}}$	$đ\mathscr{W} = d_\lambda H_{\mathbf{X}}$
Heat at trajectory level	$đ\mathscr{Q} = d_{\boldsymbol{x}} H_{\mathbf{X}}$	$đ\tilde{\mathscr{Q}} = d_{\boldsymbol{x}} H_{\mathbf{X}} + \beta \partial_\beta d_{\boldsymbol{x}} H_{\mathbf{X}} + \beta \partial_\beta d_\lambda H_{\mathbf{X}}$
1st law at trajectory level	$dH_{\mathbf{X}} = đ\mathscr{W} + đ\mathscr{Q}$	$dH_{\mathbf{X}} = đ\mathscr{W} + đ\tilde{\mathscr{Q}}$
Work at ensemble level	$đW = \int_{\boldsymbol{x}} p_{\mathbf{X}} d_\lambda H_{\mathbf{X}}$	$đW = \int_{\boldsymbol{x}} p_{\mathbf{X}} d_\lambda H_{\mathbf{X}}$
Heat at ensemble level	$đQ = \int_{\boldsymbol{x}} H_{\mathbf{X}} \, d p_{\mathbf{X}}$	$đ\tilde{Q} = \int_{\boldsymbol{x}} [(\partial_\beta \beta H_{\mathbf{X}}) d p_{\mathbf{X}} + \beta \partial_\beta (d_\lambda H_{\mathbf{X}}) p_{\mathbf{X}}]$
1st law at ensemble level	$dE_{\mathbf{X}} = đW + đQ$	$d\tilde{E}_{\mathbf{X}} = đW + đ\tilde{Q}$
2nd law (Clausius ineq.)	$dS^{\text{tot}} = dS_{\mathbf{X}} - \beta đQ = \beta(đW - dF_{\mathbf{X}}) \geqslant 0$	$dS^{\text{tot}} = d\tilde{S}_{\mathbf{X}} - \beta đ\tilde{Q} = \beta(đW - dF_{\mathbf{X}}) \geqslant 0$
Crooks FT	$p_F(W) = p_R(-W) e^{\beta(W - \Delta F_{\mathbf{X}})}$	$p_F(W) = p_R(-W) e^{\beta(W - \Delta F_{\mathbf{X}})}$
Jarzynski inequality	$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F_{\mathbf{X}}}$	$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F_{\mathbf{X}}}$

total entropy of the universe **XYZ**. Equation (4.15) then says that the total entropy production is the same, whether we calculate it using the dynamic theory of **XY** or using the reduced theory **X** alone. If we understand the dynamic theory of **X** as a consequence of coarse graining of the **XY** dynamics, then Eq. (4.15) says that entropy production is invariant under coarse graining, as long as the fast variables remain in conditional equilibrium. A similar result was obtained by Esposito [42] in the setting of master equation dynamics.

Furthermore, assuming that **XY** evolves according to Langevin dynamics (which follows if the dynamics of **Z** is much faster than that of **XY**), the Clausius inequality can be proved using the Langevin dynamics $dS_{\mathbf{XY}}[p_{\mathbf{XY}}] - \beta đQ_{\mathbf{XY}} \geqslant 0$. Hence we have

$$\begin{aligned} dS_{\mathbf{XYZ}} &= dS_{\mathbf{XY}}[p_{\mathbf{XY}}] - \beta đQ_{\mathbf{XY}} \\ &= \beta(đW_{\mathbf{XY}} - dF_{\mathbf{XY}}[p_{\mathbf{XY}}]) \geqslant 0. \end{aligned} \tag{4.16}$$

Combining Eq. (4.16) with Eq. (4.15), we finally obtain

$$\begin{aligned} dS_{\mathbf{XYZ}} &= dS_{\mathbf{X}}[p_{\mathbf{X}}] - \beta đQ \\ &= \beta(đW - dF_{\mathbf{X}}[p_{\mathbf{X}}]) \geqslant 0, \end{aligned} \tag{4.17}$$

which not only establishes the Clausius inequality but also says that the physical meaning of $dS_{\mathbf{X}}[p_{\mathbf{X}}] - \beta đQ$ is indeed the variation of total entropy of the universe.

It is interesting to rewrite Eq. (4.17) into

$$\begin{aligned} -\beta đQ &= d(S_{\mathbf{XYZ}}[p_{\mathbf{XYZ}}] - S_{\mathbf{X}}[p_{\mathbf{X}}]) \\ &= dS_{\mathbf{YZ}|\mathbf{X}}. \end{aligned} \tag{4.18}$$

Hence $-\beta đQ$ is the differential of $S_{\mathbf{YZ}|\mathbf{X}}$, the conditional Gibbs-Shannon entropy of **YZ** given the system state **X**.

V. COMPARISON WITH OTHER THEORIES

In this section, we provide a detailed comparison between this paper and several previous influential works on strong coupling thermodynamics. First of all, we list all major formulas of our theory in the center column of Table I. These formulas are identical to those of the weak coupling stochastic thermodynamic theory, with $H_{\mathbf{X}}$ understood as the Hamiltonian of mean force. In the weak coupling limit, $H_{\mathbf{X}}$ simply becomes the bare Hamiltonian of the system.

In the theory developed by Seifert [5] and critically evaluated by Talkner and Hänggi [6], the equilibrium free energy of a strongly coupled system is defined in terms of the HMF $H_{\mathbf{X}}$ as

$$F_{\mathbf{X}} = -T \ln Z_{\mathbf{X}} = -T \ln \int_{\boldsymbol{x}} e^{-\beta H_{\mathbf{X}}(\boldsymbol{x};\lambda,\beta)}, \tag{5.1}$$

which is the same as Eq. (2.17a). The equilibrium internal energy and entropy are defined as

$$\tilde{E}_{\mathbf{X}} \equiv \frac{\partial \beta F_{\mathbf{X}}}{\partial \beta} = \int p_{\mathbf{X}}^{\text{EQ}} (H_{\mathbf{X}} + \beta \, \partial_\beta H_{\mathbf{X}}), \tag{5.2a}$$

$$\tilde{S}_{\mathbf{X}} \equiv -\beta^2 \frac{\partial F_{\mathbf{X}}}{\partial \beta} = \int p_{\mathbf{X}}^{\text{EQ}} \big(-\ln p_{\mathbf{X}}^{\text{EQ}} + \beta^2 \partial_\beta H_{\mathbf{X}}\big), \tag{5.2b}$$

such that $F_{\mathbf{X}} = \tilde{E}_{\mathbf{X}} - T \tilde{S}_{\mathbf{X}}$ remains valid. (We use $\tilde{A}$ to denote the thermodynamic quantity in Seifert's theory if it is different from the corresponding quantity A in our theory.) Note that in our theory, energy and entropy are defined by Eqs. (2.17a)–(2.17d).

Reviewing the results obtained in Sec. II C, the following thermodynamic relations hold for **XY**:

$$E_{\mathbf{XY}} = \frac{\partial \beta F_{\mathbf{XY}}}{\partial \beta}, \quad S_{\mathbf{XY}} = -\beta^2 \frac{\partial F_{\mathbf{XY}}}{\partial \beta}. \tag{5.3}$$

The free energy, energy, and entropy of the bath are then defined as

$$\tilde{F}_{\mathbf{Y}} = F_{\mathbf{XY}} - F_{\mathbf{X}} = F_{\mathbf{Y}} = F_{\mathbf{Y}}^0, \tag{5.4a}$$

$$\tilde{E}_{\mathbf{Y}} = E_{\mathbf{XY}} - \tilde{E}_{\mathbf{X}}, \tag{5.4b}$$

$$\tilde{S}_{\mathbf{Y}} = S_{\mathbf{XY}} - \tilde{S}_{\mathbf{X}}, \tag{5.4c}$$

where we used Eq. (2.8b). Combining these with Eqs. (5.3), (5.2a), and (5.2b), we see that the bath energy and entropy in Seifert's theory [5] satisfy

$$\tilde{E}_{\mathbf{Y}} = \frac{\partial \beta F^0_{\mathbf{Y}}}{\partial \beta}, \quad \tilde{S}_{\mathbf{Y}} = -\beta^2 \frac{\partial F^0_{\mathbf{Y}}}{\partial \beta}, \tag{5.5}$$

where $F^0_{\mathbf{Y}}$ is the free energy of the bare bath, with the interaction switched off. These results show that in Seifert's theory, the interaction energy and correlation are completely relegated to the system. By contrast, in our theory, the interaction energy and correlation are completely relegated to the bath, if we interpret $H_{\mathbf{X}}$ as the system Hamiltonian.

Seifert [5] further bootstraps Eqs. (5.2a) and (5.2b) to the nonequilibrium case and defines fluctuating internal energy $\tilde{H}$, nonequilibrium internal energy $\tilde{E}[p_{\mathbf{X}}]$, and nonequilibrium entropy $\tilde{S}[p_{\mathbf{X}}]$ as follows:

$$\tilde{H}_{\mathbf{X}} \equiv H_{\mathbf{X}} + \beta\, \partial_\beta H_{\mathbf{X}} = \partial_\beta(\beta H_{\mathbf{X}}), \tag{5.6a}$$

$$\tilde{E}_{\mathbf{X}}[p_{\mathbf{X}}] = \int_{\boldsymbol{x}} p_{\mathbf{X}}(H_{\mathbf{X}} + \beta\, \partial_\beta H_{\mathbf{X}}), \tag{5.6b}$$

$$\tilde{S}_{\mathbf{X}}[p_{\mathbf{X}}] \equiv \int_{\boldsymbol{x}} p_{\mathbf{X}}(-\ln\ p_{\mathbf{X}} + \beta^2 \partial_\beta H_{\mathbf{X}}). \tag{5.6c}$$

The differential of entropy is then given by

$$d\tilde{S}_{\mathbf{X}} = -\int_{\boldsymbol{x}} \ln\ p_{\mathbf{X}} dp_{\mathbf{X}} + \int_{\mathbf{X}} p_{\mathbf{X}} \beta^2 \partial_\beta d_\lambda H_{\mathbf{X}}. \tag{5.7}$$

The nonequilibrium free energy is defined as

$$\begin{aligned} \tilde{F}_{\mathbf{X}}[p_{\mathbf{X}}] &\equiv \tilde{E}_{\mathbf{X}}[p_{\mathbf{X}}] - \tilde{S}_{\mathbf{X}}[p_{\mathbf{X}}] \\ &= \int_{\boldsymbol{x}} p_{\mathbf{X}}(H_{\mathbf{X}} + T\ \ln\ p_{\mathbf{X}}) \\ &= F_{\mathbf{X}}[p_{\mathbf{X}}], \end{aligned} \tag{5.8}$$

which is the same as that of our theory, Eq. (3.3).

The work at the trajectory level and the work at the ensemble level are defined in terms of change in total energy:

$$đ\tilde{W} \equiv dH_{\mathbf{XYZ}} = d_\lambda H^0_{\mathbf{X}} = d_\lambda H_{\mathbf{X}}, \tag{5.9}$$

$$đW \equiv \int_{\boldsymbol{x}} p_{\mathbf{X}}\, d_\lambda H_{\mathbf{X}}, \tag{5.10}$$

which are identical to our definitions. The heat at the trajectory level is then defined to satisfy the first law:

$$\begin{aligned} đ\tilde{Q} &\equiv d\tilde{H}_{\mathbf{X}} - đ\tilde{W} \\ &= d_{\boldsymbol{x}} H_{\mathbf{X}} + \beta \partial_\beta d_{\boldsymbol{x}} H_{\mathbf{X}} + \beta \partial_\beta d_\lambda H_{\mathbf{X}}, \end{aligned} \tag{5.11}$$

$$đ\tilde{Q} \equiv \int_{\boldsymbol{x}} [(\partial_\beta \beta H_{\mathbf{X}}) dp_{\mathbf{X}} + \beta \partial_\beta (d_\lambda H_{\mathbf{X}}) p_{\mathbf{X}}]. \tag{5.12}$$

The left-hand side of the Clausius inequality can be calculated:

$$\begin{aligned} d\tilde{S}_{\mathbf{X}} - \beta đ\tilde{Q} = dS_{\mathbf{X}} - \beta đQ &= \beta(đW - dF_{\mathbf{X}}) \\ &= -\int_{\boldsymbol{x}} (\ln\ p_{\mathbf{X}} + \beta H_{\mathbf{X}}) dp_{\mathbf{X}}, \end{aligned} \tag{5.13}$$

which is again the same as in our theory. As a consequence, the first and second laws of thermodynamics in Seifert's theory [5] are equivalent to those in our theory. This means that these two theories are equivalent to each other, even though they use different definitions of internal energy, entropy, and heat. Major formulas of Seifert's theory are displayed in the right column of Table I.

Talkner and Hänggi [6,8] accept the definitions of equilibrium thermodynamic quantities, Eqs. (5.2a) and (5.2b). Yet they argue that the nonequilibrium thermodynamic quantities cannot be uniquely determined from their equilibrium versions, which is of course valid. They also argue that the Hamiltonian of mean force cannot be uniquely determined from the equilibrium distribution of system variables alone [43]. They further discuss more serious ambiguities associated with the definition of nonequilibrium work for quantum systems.

Jarzynski [7] develops a more comprehensive (and hence more complex) theory for strong coupling thermodynamics and systematically discusses the definitions of internal energy, entropy, volume, pressure, enthalpy, and Gibbs free energy. Using a pebble immersed in a liquid as a metaphor, Jarzynski establishes his formalism around the concept of volume, whose definition is somewhat arbitrary. All other thermodynamic variables are uniquely fixed by thermodynamic consistency once the system volume is (arbitrarily) defined. Jarzynski further shows that Seifert's theory [5] is a special case of his (Jarzynski's) framework, i.e., the "partial molar representation." Jarzynski discusses in great detail the "bare representation," where the system enthalpy coincides with the HMF. The total entropy production is, however, the same in both representations. The heat and work in the bare representation are formally identical to those in our theory. We note that for many small systems, volume or pressure is seldom controlled. It is then unnecessary to distinguish energy from enthalpy, or Helmholtz free energy from Gibbs free energy.

In all the works discussed above, the interaction Hamiltonian H_I is assumed to be independent of the external parameter λ, whereas time-scale separation is not assumed. As a consequence, it is possible to prove the integrated Clausius inequality $\Delta S - \beta Q \geqslant 0$ for a finite process but not possible to prove the differential Clausius inequality $dS - \beta đQ \geqslant 0$ for every infinitesimal evolution step in the process. Barring the issues of TSS and of λ dependence of the interaction Hamiltonian H_I, our theory can be understood as a simplification of Jarzynski's bare representation, with the HMF and free energy playing the role of enthalpy and Gibbs free energy.

Strasberg and Esposito [14] studied the consequences of TSS in the settings both of master equation theory and of Hamiltonian dynamics. For master equation theory, using the conditional equilibrium nature of the fast variables, they show that a reduced theory of slow variables can be derived once the fast variables are averaged out. Note, however, that the heat and internal energy as given by Eqs. (33)–(35) of Ref. [14] still pertain to the combined system. To obtain a thermodynamic theory for the slow variables alone, one would have to subtract from these quantities the parts due to fast variables. The

resulting quantities would then pertain to the slow variables only, and remain finite even if the dimension of the fast variables goes to infinity. For Hamiltonian dynamics, Strasberg and Esposito *propose* a definition of total entropy production as the relative entropy and show that, with TSS, it is equivalent to that in Seifert's theory [5], which is also equivalent to entropy production in our theory, as we have demonstrated in Eq. (5.13). By this, they confirm the consistency of Seifert's strong coupling theory.

By contrast, in this paper, we use TSS to carry out a different decomposition of the Hamiltonian as discussed in Sec. II A. This leads to a remarkable situation where all formulas of the weak coupling theory of stochastic thermodynamics remain applicable even in the strong coupling regime. These formulas are significantly simpler than those in Seifert's strong coupling theory [5]. For a comparison, see Table I.

The differences between the present theory and Seifert's theory [5] are, however, not completely notational. Consider a "fast" slow process with time duration dt where λ changes by $d\lambda$. It is slow enough that the bath remains in conditional equilibrium, and our stochastic thermodynamic theory remains applicable. Yet it is also fast enough that the distribution $p_{\mathbf{X}}$ barely changes. Such a process can always be realized if TSS is satisfied. Hence we have $d_\lambda H_{\mathbf{X}} \neq 0$, but $dp_{\mathbf{X}} = 0$. According to the present theory, then both $dS_{\mathbf{X}}$ and $đQ$ vanish, and hence the variation of total entropy $dS_{\mathbf{X}} - \beta đQ$ also vanishes. Now in Seifert's theory, $d\tilde{S}_{\mathbf{X}}$ and $đ\tilde{Q}$ are given by Eqs. (5.7) and (5.12), respectively. Neither of these two vanishes even if $dp_{\mathbf{X}} = 0$; yet the variation of the total entropy $d\tilde{S}_{\mathbf{X}} - \beta đ\tilde{Q}$ does vanish. This means that in Seifert's theory there is an exchange of entropy between the system and the bath even though $p_{\mathbf{X}}$ remains unchanged. While this does not violate the second law of thermodynamics, it does contradict the common intuition about entropy as a measure of a multitude of system states: It is very strange if the pdf of a system variable stays unchanged and yet the system entropy changes suddenly! From this perspective, the present theory is more natural and intuitive.

VI. CONCLUSION

In this paper, we have demonstrated that the usual theory of strong coupling thermodynamics and stochastic thermodynamics, which is based on the assumption of weak coupling between the system and its environment, can be made applicable in the strong coupling regime, if we define the Hamiltonian of mean force as the system Hamiltonian. Our result is consistent with previous theories by various authors, in the sense that the first and second laws in different theories are mathematically equivalent. Overall, the present work can be understood as a reinterpretation, synthesis, and simplification of various previous theories of strong coupling stochastic thermodynamics.

In future work, we will conduct a systematic study of the coarse-graining process, i.e., integrating out fast variables to obtain an effective dynamic theory for slow variables, with the ratio of time scales between the slow and fast variables treated as a small parameter. If this ratio is small but nonzero, there should be a slight deviation of fast variables from conditional equilibrium. We shall analyze how this deviation leads to modification of dissipation in the dynamics of slow variables. We shall also extend our theory to the quantum case and develop a thermodynamic theory for small open quantum systems strongly coupled to the environment.

ACKNOWLEDGMENTS

X.X. acknowledges support from NSFC Grant No. 11674217 as well as Shanghai Municipal Science and Technology Major Project (Grant No. 2019SHZDZX01). Z.T. acknowledges support from NSFC Grant No. 11675017. X.X. is also thankful for additional support from the Shanghai Talent Program.

[1] U. Seifert, Entropy Production along a Stochastic Trajectory and an Integral Fluctuation Theorem, Phys. Rev. Lett. **95**, 040602 (2005).

[2] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, Rep. Prog. Phys. **75**, 126001 (2012).

[3] C. Jarzynski, Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale, Annu. Rev. Condens. Matter Phys. **2**, 329 (2011).

[4] In general, we calculate these differentials up to first order in dt, the differential time. If $\lambda(t)$ and $\boldsymbol{x}(t)$ are both differentiable, we have $d_\lambda H(\boldsymbol{x}, \lambda) = \frac{\partial H}{\partial \lambda} d\lambda$ and $d_{\boldsymbol{x}} H(\boldsymbol{x}, \lambda) = \frac{\partial H}{\partial \boldsymbol{x}} d\boldsymbol{x}$. If $\boldsymbol{x}(t)$ is nondifferentiable, which happens if $\boldsymbol{x}(t)$ is a typical dynamic trajectory of Brownian motion, then we may need to expand $d_{\boldsymbol{x}} H(\boldsymbol{x}, \lambda)$ up to the second order in $d\boldsymbol{x}$. In any case, the cross term which is proportional to $d\boldsymbol{x} d\lambda$ is always negligible.

[5] U. Seifert, First and Second Law of Thermodynamics at Strong Coupling, Phys. Rev. Lett. **116**, 020601 (2016).

[6] P. Talkner and P. Hänggi, Open system trajectories specify fluctuating work but not heat, Phys. Rev. E **94**, 022143 (2016).

[7] C. Jarzynski, Stochastic and Macroscopic Thermodynamics of Strongly Coupled Systems, Phys. Rev. X **7**, 011008 (2017).

[8] P. Talkner and P. Hänggi, Colloquium: Statistical mechanics and thermodynamics at strong coupling: Quantum and classical, Rev. Mod. Phys. **92**, 041002 (2020).

[9] P. Strasberg, G. Schaller, N. Lambert, and T. Brandes, Non equilibrium thermodynamics in the strong coupling and non-Markovian regime based on a reaction coordinate mapping, New J. Phys. **18**, 073007 (2016).

[10] E. Aurell, Unified picture of strong-coupling stochastic thermodynamics and time reversals, Phys. Rev. E **97**, 042112 (2018).

[11] H. J. D. Miller and J. Anders, Entropy production and time asymmetry in the presence of strong interactions, Phys. Rev. E **95**, 062123 (2017).

[12] M. F. Gelin and M. Thoss, Thermodynamics of a sub-ensemble of a canonical ensemble, Phys. Rev. E **79**, 051121 (2009).

[13] R. de Miguel and J. M. Rubí, Strong coupling and nonextensive thermodynamics, Entropy **22**, 975 (2020).

[14] P. Strasberg and M. Esposito, Stochastic thermodynamics in the strong coupling regime: An unambiguous approach based on coarse graining, Phys. Rev. E **95**, 062101 (2017).

[15] P. Strasberg and M. Esposito, Measurability of nonequilibrium thermodynamics in terms of the Hamiltonian of mean force, Phys. Rev. E **101**, 050101(R) (2020).

[16] M. Campisi, P. Talkner, and P. Hänggi, Fluctuation Theorem for Arbitrary Open Quantum Systems, Phys. Rev. Lett. **102**, 210401 (2009).

[17] J.-T. Hsiang and B.-L. Hu, Zeroth law in quantum thermodynamics at strong coupling: In equilibrium, not at equal temperature, Phys. Rev. D **103**, 085004 (2021).

[18] W.-M. Huang and W.-M. Zhang, Strong coupling quantum thermodynamics with renormalized Hamiltonian and temperature, arXiv:2010.01828.

[19] M. Carrega, P. Solinas, M. Sassetti, and U. Weiss, Energy Exchange in Driven Open Quantum Systems at Strong Coupling, Phys. Rev. Lett. **116**, 240403 (2016).

[20] A. Rivas, Strong Coupling Thermodynamics of Open Quantum Systems, Phys. Rev. Lett. **124**, 160601 (2020).

[21] M. Perarnau-Llobet, H. Wilming, A. Riera, R. Gallego, and J. Eisert, Strong Coupling Corrections in Quantum Thermodynamics, Phys. Rev. Lett. **120**, 120602 (2018).

[22] P. Strasberg, Repeated Interactions and Quantum Stochastic Thermodynamics at Strong Coupling, Phys. Rev. Lett. **123**, 180604 (2019).

[23] L. Pucci, M. Esposito, and L. Peliti, Entropy production in quantum Brownian motion, J. Stat. Mech.: Theory Exp. (2013)P04005.

[24] M. Esposito, M. A. Ochoa, and M. Galperin, Nature of heat in strongly coupled open quantum systems, Phys. Rev. B **92**, 235440 (2015).

[25] S. Abe and A. K. Rajagopal, Validity of the Second Law in Nonextensive Quantum Thermodynamics, Phys. Rev. Lett. **91**, 120601 (2003).

[26] K. Ptaszyński and M. Esposito, Entropy Production in Open Systems: The Predominant Role of Intraenvironment Correlations, Phys. Rev. Lett. **123**, 200603 (2019).

[27] M. Esposito, K. Lindenberg, and C. Van den Broeck, Entropy production as correlation between system and reservoir, New J. Phys. **12**, 013013 (2010).

[28] R. de Miguel and J. M. Rubí, Statistical mechanics at strong coupling: A bridge between Landsberg's energy levels and Hill's nanothermodynamics, Nanomaterials **10**, 2471 (2020).

[29] B. Roux and T. Simonson, Implicit solvent models, Biophys. Chem. **78**, 1 (1999).

[30] C. Jarzynski, Nonequilibrium work theorem for a system strongly coupled to a thermal environment, J. Stat. Mech.: Theory Exp. (2004)P09005.

[31] J. G. Kirkwood, Statistical mechanics of fluid mixtures, J. Chem. Phys. **3**, 300 (1935).

[32] M. Born and V. Fock, Proof of the adiabatic theorem, Z. Phys. **51**, 165 (1928).

[33] M. Born and J. R. Oppenheimer, Zur Quantentheorie der Molekeln [On the quantum theory of molecules], Ann. Phys. (Berlin) **389**, 457 (1927).

[34] C. W. Gardiner, *Handbook of Stochastic Methods*, 3rd ed. (Springer, Berlin, 2004).

[35] L. Michaelis and M. Menten, Die Kinetik der Invertinwirkung, Biochem. Z. **49**, 333 (1913).

[36] G. Pavliotis and A. Stuart, *Multiscale Methods: Averaging and Homogenization* (Springer, New York, 2008).

[37] M. Ding, Z. Tu, and X. Xing, Covariant formulation of nonlinear Langevin theory with multiplicative Gaussian white noises, Phys. Rev. Res. **2**, 033381 (2020).

[38] M. Ding and X. Xing, Covariant non-equilibrium thermodynamics from Ito-Langevin dynamics, arXiv:2105.14534.

[39] Recall that we are dealing with classical systems, where there is no problem of noncommutativity. For quantum systems, Eq. (2.2) is equal to the Hamiltonian of mean force only if $H^0_{\mathbf{X}}$ commutes with $H_{\mathbf{Y}} + H^0_I$.

[40] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 2012).

[41] This is reasonable as long as we do not take the fast time scale to zero.

[42] M. Esposito, Stochastic thermodynamics under coarse graining, Phys. Rev. E **85**, 041125 (2012).

[43] See, however, the recent work by Strasberg and Esposito on this issue [15].