Stochastic thermodynamics in the strong coupling regime: An unambiguous approach based on coarse graining

Philipp Strasberg[*] and Massimiliano Esposito

Complex Systems and Statistical Mechanics, Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg, Luxembourg

(Received 15 March 2017; published 1 June 2017)

We consider a classical and possibly driven composite system $X \otimes Y$ weakly coupled to a Markovian thermal reservoir R so that an unambiguous stochastic thermodynamics ensues for $X \otimes Y$. This setup can be equivalently seen as a system X strongly coupled to a non-Markovian reservoir $Y \otimes R$. We demonstrate that only in the limit where the dynamics of Y is much faster than X, our unambiguous expressions for thermodynamic quantities, such as heat, entropy, or internal energy, are equivalent to the strong coupling expressions recently obtained in the literature using the Hamiltonian of mean force. By doing so, we also significantly extend these results by formulating them at the level of instantaneous rates and by allowing for time-dependent couplings between X and its environment. Away from the limit where Y evolves much faster than X, previous approaches fail to reproduce the correct results from the original unambiguous formulation, as we illustrate numerically for an underdamped Brownian particle coupled strongly to a non-Markovian reservoir.

I. INTRODUCTION

Establishing the laws of thermodynamics for a given setup is not only beneficial for practical purposes but also provides an important consistency check for the validity of the model and provides much deeper insights into the structure of the problem. Yet, establishing these laws for small-scale systems away from the well-established weak-coupling and Markovian limit can be very challenging.

This paper focuses on the case of a small, driven classical system in strong contact with a single environment. This case has attracted a lot of attention recently and was mostly tackled by introducing a *Hamiltonian of mean force* (HMF), classically [1–7] as well as quantum-mechanically [2,5,8–11]. However, the question of what exactly are the correct definitions for heat, internal energy, and other quantities causes already controversies at the classical level [2–4,6].

We present an enlightening perspective on this problem by considering two coupled systems $X \otimes Y$, which are in weak contact with a large thermal reservoir R and obey standard stochastic thermodynamics. By realizing that the system X can be strongly coupled to the system Y, we see that the situation is *equivalent* to a system X in strong contact with an environment $E = Y \otimes R$. In fact, for many relevant scenarios it makes sense that the system X only couples strongly to a subpart $Y \subset E$ of the environment but not to each degree of freedom of E. This picture is also supported by our example at the end of the paper.

The benefit of our approach is that we start from a well-defined thermodynamics with unambiguous definitions and we can then compare under which conditions previous approaches based on the HMF coincide with them. Our framework can be seen as an application of the laws of thermodynamics under coarse graining as detailed in Ref. [12], also see Refs. [13,14].

Outline. We start by presenting the thermodynamic description of the combined system $X \otimes Y$ in contact with R in Sec. II. We show what changes if we coarse-grain Y and

consider the important limit where Y evolves much faster than X such that it can be adiabatically eliminated and a closed thermodynamic description for X alone emerges. In Sec. III, we turn the situation around and start from a description of X coupled to E and derive an exact inequality. Based on this, we recapitulate the thermodynamics previously established using the HMF and we show that we are able to rederive and greatly extend these results if Y evolves much faster than X. Beyond that we explicitly quantify the difference in the two proposed definitions of entropy production and we use an example in Sec. IV to illustrate generic features in the thermodynamics of strongly coupled systems, demonstrating that knowledge of the HMF alone does not suffice to reproduce the original thermodynamics. It also provides a strategy to identify a system Y if the initial setup is described at the level of $X \otimes E$. Section V summarizes the main implications of our findings.

II. SETUP

A. Basic quantities

A sketch of the setup is shown in Fig. 1. Two systems X and Y interact with each other and their joint Hamiltonian is assumed to be of the form

$$E_{xy}(\lambda_t) = E_x(\lambda_t) + V_{xy}(\lambda_t) + E_y. \tag{1}$$

Here, the energy $E_x(\lambda_t)$ of system X as well as the interaction energy $V_{xy}(\lambda_t)$ can be time-dependent due to some externally controlled parameters λ_t. The energy E_y of system Y is assumed to be time-independent.

Since the joint system $X \otimes Y$ is weakly coupled to a large thermal reservoir at inverse temperature β, we assume that its dynamics can be modeled by a Markovian master equation (ME) of the form

$$d_t p_{xy}(t) = \sum_{x',y'} R_{xy,x'y'}(\lambda_t) p_{x'y'}(t). \tag{2}$$

Here, $p_{xy}(t)$ denotes the probability to find the system in state xy at time t. The rate matrix obeys $\sum_{x,y} R_{xy,x'y'}(\lambda_t) = 0$,
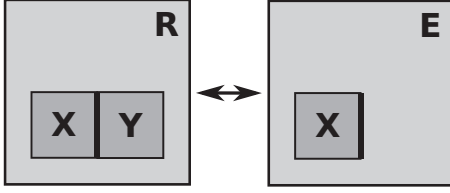
[*]philipp.strasberg@uni.lu

FIG. 1. Sketch of the setup: two systems X and Y weakly interact with a large thermal reservoir R. Equivalently, the system X interacts strongly with a composite environment $E \equiv Y \otimes R$.

which ensures conservation of probability [$\sum_{x,y} p_{xy}(t) = 1$]. Furthermore, we assume local detailed balance

$$\frac{R_{xy,x'y'}(\lambda_t)}{R_{x'y',xy}(\lambda_t)} = e^{-\beta[E_{xy}(\lambda_t) - E_{x'y'}(\lambda_t)]}, \quad (3)$$

which allows a physical interpretation of the ME and especially a consistent thermodynamic description. The rest of this section then follows standard stochastic thermodynamics [15–17].

For this purpose, we introduce the following quantities:

$$U_{XY}(t) \equiv \langle E_{xy}(\lambda_t) \rangle \quad \text{(internal energy)}, \quad (4)$$

$$\dot{W}(t) \equiv \langle d_t E_{xy}(\lambda_t) \rangle \quad \text{(work rate)}, \quad (5)$$

$$\dot{Q}(t) \equiv \sum_{x,y} E_{xy}(\lambda_t) d_t p_{xy}(t) \quad \text{(heat rate)}, \quad (6)$$

$$S_{XY}(t) \equiv \langle -\ln p_{xy}(t) \rangle \quad \text{(entropy)}. \quad (7)$$

Here, we have denoted the ensemble average with respect to any solution $p_{xy}(t)$ of the ME above by $\langle f_{xy}(t) \rangle \equiv \sum_{x,y} p_{xy}(t) f_{xy}(t)$. Furthermore, the thermodynamic entropy S coincides in the weak coupling regime with the definition of Shannon entropy.

Based on the definitions above, it is straightforward to derive the first law:

$$d_t U_{XY}(t) = \dot{W}(t) + \dot{Q}(t) \quad (8)$$

(we define heat and work positive if they increase the energy of $X \otimes Y$). Furthermore, the second law states that the overall entropy production rate is positive:

$$\dot{\Sigma}(t) \equiv d_t S_{XY}(t) - \beta \dot{Q}(t) \geq 0. \quad (9)$$

Its positivity can be proven by noting the identity

$$\dot{\Sigma}(t) = \sum_{x,x',y,y'} R_{xy,x'y'}(\lambda_t) p_{x'y'}(t)$$
$$\times \ln \frac{R_{xy,x'y'}(\lambda_t) p_{x'y'}(t)}{R_{x'y',xy}(\lambda_t) p_{xy}(t)} \quad (10)$$

and using $-\ln x \geq 1 - x$. Another useful identity is

$$\dot{\Sigma}(t) = -\partial_t |_{\lambda(t)} D[p_{xy}(t) \| p_{xy}^{\text{eq}}(\lambda_t)] \geq 0, \quad (11)$$

where the partial derivative $\partial_t |_{\lambda(t)}$ indicates that the change of $D[p_{xy}(t) \| p_{xy}^{\text{eq}}(\lambda_t)]$ is evaluated at fixed λ_t. Here, we introduced the equilibrium (Gibbs, thermal) state,

$$p_{xy}^{\text{eq}}(\lambda_t) \equiv \frac{e^{-\beta E_{xy}(\lambda_t)}}{\mathcal{Z}_{XY}(\lambda_t)}, \quad (12)$$

which depends parametrically on time, and $\mathcal{Z}_{XY}(\lambda_t)$ denotes the equilibrium partition function. Also, the concept of relative entropy,

$$D[p_x \| q_x] \equiv \sum_x p_x (\ln p_x - \ln q_x) \geq 0, \quad (13)$$

which is always positive for any two probability distributions p_x and q_x, will be used later on.

Finally, let us introduce the concept of a nonequilibrium free energy $F_{XY}(t) = U_{XY}(t) - \beta^{-1} S_{XY}(t)$, which is defined for any state $p_{xy}(t)$. Using this, we can reformulate the second law as

$$\dot{\Sigma}(t) = \beta[\dot{W}(t) - d_t F_{XY}(t)] \geq 0. \quad (14)$$

Whenever a system α (where α could stand for $XY, X, E, \dots$ depending on the situation) is at equilibrium, it is useful to note the relations

$$\mathcal{F}_\alpha(\lambda_t) = -\beta^{-1} \ln \mathcal{Z}_\alpha(\lambda_t), \quad (15)$$

$$\mathcal{U}_\alpha(\lambda_t) = \partial_\beta [\beta \mathcal{F}_\alpha(\lambda_t)], \quad (16)$$

$$\mathcal{S}_\alpha(\lambda_t) = \beta^2 \partial_\beta \mathcal{F}_\alpha(\lambda_t), \quad (17)$$

for the equilibrium free energy, internal energy, and entropy, respectively. Note that we use calligraphic letters $\mathcal{F}, \mathcal{Z}, \mathcal{U}$, and $\mathcal{S}$ to denote thermodynamic quantities at equilibrium.

Below, to keep a compact notation, we will often omit the dependence on λ_t in the notation.

B. Coarse-graining

We now shift our attention to system X alone and mostly follow Ref. [12] for the rest of this section. For this purpose we split the joint probability into a conditional and marginal probability as

$$p_{xy}(t) = p_{y|x}(t) p_x(t), \quad (18)$$

with $p_x(t) = \sum_y p_{xy}(t)$ and $\sum_y p_{y|x}(t) = 1$. It is not hard to deduce that $p_x(t)$ evolves according to the ME,

$$d_t p_x(t) = \sum_{x'} R_{x,x'} p_{x'}(t), \quad (19)$$

where the new effective rate matrix is given by

$$R_{x,x'} = R_{x,x'}(\lambda_t, t) = \sum_{y,y'} R_{xy,x'y'}(\lambda_t) p_{y'|x'}(t). \quad (20)$$

In general, it depends explicitly on time due to the time-dependence of $p_{y'|x'}(t)$ and solving Eq. (19) is equally hard as solving the original ME unless further assumptions are made.

Nevertheless, there is an apparent second law related to the reduced dynamics of X [12]:

$$\dot{\Sigma}^{(1)}(t) = \sum_{x,x'} R_{x,x'} p_{x'}(t) \ln \frac{R_{x,x'} p_{x'}(t)}{R_{x',x} p_x(t)} \geq 0, \quad (21)$$

which can be rewritten as

$$\dot{\Sigma}^{(1)}(t) = d_t S_{XY}(t) - \beta \dot{Q}^{(1)}(t) \geq 0. \quad (22)$$

Here, we introduced the apparent heat flow

$$\dot{Q}^{(1)}(t) \equiv -\frac{1}{\beta} \sum_{x,x'} R_{x,x'} p_{x'}(t) \ln \frac{R_{x,x'}}{R_{x',x}} + \frac{1}{\beta} d_t S_{Y|X}(t), \quad (23)$$

and $S_{Y|X}$ denotes the conditional Shannon entropy

$$S_{Y|X}(t) = -\sum_x p_x(t) \sum_y p_{y|x}(t) \ln p_{y|x}(t), \quad (24)$$

which fulfills $S_{XY} = S_X + S_{Y|X}$ with $S_X \equiv -\sum_x p_x(t) \ln p_x(t)$. Unfortunately, at this general level there is no relation between $\dot{Q}^{(1)}$ and the real heat flow $\dot{Q}$, making it hard to establish a local version of the first law. Furthermore, note that $\dot{\Sigma}^{(1)}$ always underestimates the true entropy production $\dot{\Sigma} \geqslant \dot{\Sigma}^{(1)} \geqslant 0$.

C. Time-scale separation

There is an important limit, in which Y evolves much faster than X and can be adiabatically eliminated. We will refer to this as time-scale separation (TSS). Within TSS, one assumes that

$$R_{xy,xy'} \gg R_{xy,x'y'} \quad (25)$$

for $x \neq x'$. To get a simple description, we also assume that for each given x all states y are connected, i.e., $R_{xy,xy'} \neq 0$ for all y, y'.

Under these conditions, one can show that the conditional probabilities $p_{y|x}(t)$ equilibrate and can be written as[1]

$$\bar{p}_{y|x} = \bar{p}_{y|x}(\lambda_t) = e^{-\beta(E_{xy} - F_{Y|x})}. \quad (26)$$

Normalization is ensured by choosing

$$F_{Y|x} \equiv -\beta^{-1} \ln \sum_y e^{-\beta E_{xy}(\lambda_t)}$$

$$= E_x - \beta^{-1} \ln \langle e^{-\beta V_{xy}} \rangle_Y^{\text{eq}} + \mathcal{F}_Y, \quad (27)$$

where $\langle \dots \rangle_Y^{\text{eq}}$ denotes the ensemble average with respect to $p_y^{\text{eq}} = e^{-\beta E_y}/\mathcal{Z}_Y$. Note that $F_{Y|x} = F_{Y|x}(\lambda_t)$ depends parametrically on time. In contrast, the equilibrium free energy $\mathcal{F}_Y$ has no time-dependence. Although it appears in the definition of $F_{Y|x}$, we remark that the reduced state of Y is not given by the equilibrium state p_y^{eq}.

Within TSS we denote the rate matrix $R_{x,x'}$ by $\bar{R}_{x,x'} = \bar{R}_{x,x'}(\lambda_t)$, which now depends only parametrically on time and greatly simplifies the solution of Eq. (19). Furthermore, it fulfills an effective local detailed balance relation of the form

$$\frac{\bar{R}_{x,x'}}{\bar{R}_{x',x}} = e^{-\beta(F_{Y|x} - F_{Y|x'})}, \quad (28)$$

which makes the meaning of $F_{Y|x}$ as a free-energy landscape for system X transparent. Using this, we can express the

apparent heat flow Eq. (23) as

$$\dot{Q}^{(1)}(t) = \sum_{x,x'} F_{Y|x} \bar{R}_{x,x'} p_{x'}(t) + \frac{1}{\beta} d_t S_{Y|X}(t), \quad (29)$$

which now coincides with the real heat flow:

$$\dot{Q}(t) = \dot{Q}^{(1)}(t). \quad (30)$$

To prove Eq. (30), it is useful to note that

$$\frac{1}{\beta} d_t S_{Y|X}(t) = -\frac{1}{\beta} \sum_{x,y} [d_t p_x(t) \bar{p}_{y|x}(\lambda_t)] \ln \bar{p}_{y|x}(\lambda_t)$$

$$= \sum_{x,y} (E_{xy} - F_{Y|x})[d_t p_x(t) \bar{p}_{y|x}(\lambda_t)], \quad (31)$$

where we used Eq. (26). Then, after realizing that $\sum_y \bar{p}_{y|x}(\lambda_t) = 1$ and $\sum_y d_t \bar{p}_{y|x}(\lambda_t) = 0$, we get

$$\frac{1}{\beta} d_t S_{Y|X}(t) = \sum_{x,y} E_{xy} d_t [\bar{p}_{y|x}(\lambda_t) p_x(t)] + \sum_x F_{Y|x} d_t p_x(t). \quad (32)$$

Plugging this result into Eq. (29), we finally obtain

$$\dot{Q}^{(1)}(t) = \sum_{x,y} E_{xy} d_t [\bar{p}_{y|x}(\lambda_t) p_x(t)], \quad (33)$$

which equals our original definition Eq. (6) within TSS.

Furthermore, it makes sense to rewrite the internal energy as $U_{XY}(t) = \sum_x U_{XY|x} p_x(t)$ where we introduced the average internal energy *conditioned* on the state x:

$$U_{XY|x} = U_{XY|x}(\lambda_t) \equiv \sum_y E_{xy} \bar{p}_{y|x}. \quad (34)$$

Formally, the first law remains the same as before,

$$d_t U_{XY}(t) = \dot{W}(t) + \dot{Q}(t) = \dot{W}(t) + \dot{Q}^{(1)}(t). \quad (35)$$

In contrast to the general case, however, the time-dependence of all quantities comes only from the dynamical time-dependence of the system X alone and the parametric dependence on λ_t. The same observation holds true for second law of thermodynamics, which can be expressed as

$$\dot{\Sigma}(t) = d_t S_{XY}(t) - \beta \dot{Q}^{(1)}(t) \geqslant 0. \quad (36)$$

Thus, within TSS we have indeed $\dot{\Sigma} = \dot{\Sigma}^{(1)}$. For later purposes it will be also convenient to note the following two identities:

$$U_{XY|x} = \partial_\beta(\beta F_{Y|x}), \quad (37)$$

$$S_{Y|x} = -\sum_y \bar{p}_{y|x} \ln \bar{p}_{y|x} = \beta^2 \partial_\beta F_{Y|x}, \quad (38)$$

which look remarkably similar to Eqs. (16) and (17). Proving them follows from straightforward though tedious algebraic manipulations, which we will not display here.

Finally, we briefly mention how to extend the results above to the stochastic level following the well-established procedure [16,17]. This will also underline the fact that any information about Y enters only statically in the description. If the system starts at time t_0 in state x_0, jumps at time $t_1 > t_0$ to x_1, and stays in that state until it jumps at time $t_2 > t_1$ to x_2, etc., we denote

---

[1]Of course, there are also alternative parametrizations possible. For instance, $\bar{p}_{y|x} = e^{-\beta[E_y + V_{xy}(\lambda_t) - F'_{Y|x}(\lambda_t)]}$ with $F'_{Y|x}(\lambda_t) = F_{Y|x}(\lambda_t) - E_x(\lambda_t)$. This does not affect the resulting thermodynamics at the end.

this trajectory by $\mathbf{x}_t \equiv (x_0, t_0; x_1, t_1; \dots)$. Then, the fluctuating internal energy at each instant t is given by

$$U_{XY|x_t} = \sum_y E_{x_t y} \bar{p}_{y|x_t}. \tag{39}$$

The work along the trajectory $\mathbf{x}_t$ becomes

$$W[\mathbf{x}_t] = \int_{t_0}^{t} ds \dot{\lambda}_s \sum_y \bar{p}_{y|\mathbf{x}_s} \partial_{\lambda_s} E_{\mathbf{x}_s y}, \tag{40}$$

and the stochastic entropy is defined as

$$S_{XY}[x_t] = -\ln p_{x_t}(t) - \sum_y p_{y|x_t}(\lambda_t) \ln p_{y|x_t}(\lambda_t). \tag{41}$$

Finally, the heat can be decomposed as

$$\beta Q[\mathbf{x}_t] = \sum_j \ln \frac{\bar{R}_{x_j, x_{j+1}}}{\bar{R}_{x_{j+1}, x_j}} - \sum_y \bar{p}_{y|x_t}(\lambda_t) \ln \bar{p}_{y|x_t}(\lambda_t)$$
$$+ \sum_y \bar{p}_{y|x_0}(\lambda_0) \ln \bar{p}_{y|x_0}(\lambda_0), \tag{42}$$

where the sum indexed by j runs over all jumps, which have happened from t_0 to t. Since $p_x(t)$ obeys a Markovian ME with rates that fulfill the local detailed balance relation Eq. (28), it is clear that the integral and detailed fluctuation theorems are also obeyed, e.g.,

$$\langle\langle e^{-\Sigma[\mathbf{x}_t]} \rangle\rangle = 1, \tag{43}$$

where $\Sigma[\mathbf{x}_t] = S_{XY}[x_t] - S_{XY}[x_0] - \beta Q[\mathbf{x}_t]$ and $\langle\langle \dots \rangle\rangle$ denotes an ensemble average over all trajectories $\mathbf{x}_t$.

III. THE HAMILTONIAN OF MEAN FORCE

A. Exact identities

In this section, we turn the situation around and consider a system X in contact with an environment E as shown on the right-hand side of Fig. 1 and we only use in Sec. III B the decomposition $E = Y \otimes R$. We assume that the combined system $X \otimes E$ is isolated and obeys Hamiltonian dynamics with Hamiltonian

$$E_{xe}(\lambda_t) = E_x(\lambda_t) + V_{xe}(\lambda_t) + E_e, \tag{44}$$

where e denotes a microstate of the environment. Note that we will, in general, denote thermodynamic quantities in this section by a "tilde" to distinguish them from previously introduced quantities. Their relation will be clarified in Sec. III B.

As in Refs. [3,4], we assume that the initial state of $X \otimes E$ reads

$$p_{xe}(0) = p_x(0)\bar{p}_{e|x}(\lambda_0) = p_x(0)e^{-\beta(E_{xe} - F_{E|x})}, \tag{45}$$

where $p_x(0)$ is an arbitrary initial system state and $\bar{p}_{e|x}(\lambda_0)$ denotes the equilibrium state of E conditioned on a microstate x of the system. Clearly, $F_{E|x}$ is defined as in Eq. (27) with Y replaced by E. The state $\bar{p}_{e|x}$ can be more elegantly expressed by introducing the HMF,

$$E_x^*(\lambda_t) \equiv E_x(\lambda_t) - \frac{1}{\beta} \ln \langle e^{-\beta V_{xe}(\lambda_t)} \rangle_E^{\text{eq}}, \tag{46}$$

which has been successfully used for a long time in thermostatics [18,19]. Using this, we find

$$\bar{p}_{e|x}(\lambda_t) = \frac{e^{-\beta[E_{xe}(\lambda_t) - E_x^*(\lambda_t)]}}{\mathcal{Z}_E} \tag{47}$$

and also the important relation

$$E_x^* = F_{E|x} - \mathcal{F}_E. \tag{48}$$

Given the initial state Eq. (45), we follow Ref. [20] and define an entropy production

$$\tilde{\Sigma}(t) \equiv D[p_{xe}(t) \| p_x(t)\bar{p}_{e|x}(\lambda_t)] \geqslant 0, \tag{49}$$

which measures the deviation of the true state $p_{xe}(t)$ from an idealized reference state $p_x(t)\bar{p}_{e|x}(\lambda_t)$. Note that definition Eq. (49) differs from Ref. [20] only in the choice of the reference state. We discuss in Appendix A how both are related. We will now show that Eq. (49) coincides with the definition used in Ref. [3] as was independently and simultaneously noted in Ref. [7].

It is convenient to rewrite Eq. (49) as

$$\tilde{\Sigma}(t) = \Delta S_X(t) - \sum_{x,e} [p_{xe}(t) \ln \bar{p}_{e|x}(\lambda_t) - p_{xe}(0) \ln \bar{p}_{e|x}(\lambda_0)], \tag{50}$$

where we used that the Shannon entropy of the global system $X \otimes E$ remains constant under Hamiltonian dynamics, $S_{XE}(t) = S_{XE}(0)$. Furthermore, we use the notation $\Delta f(t) \equiv f(t) - f(0)$ for any time-dependent function $f(t)$. Now, in accordance with phenomenological nonequilibrium thermodynamics, we would like to split $\tilde{\Sigma}(t)$ into two parts:

$$\tilde{\Sigma}(t) = \Delta \tilde{S}_X(t) - \beta \tilde{Q}(t) \geqslant 0. \tag{51}$$

Without additional information, there is obviously no unique splitting of these two quantities at this formal level, which essentially translates the results of Ref. [4] into our framework. For the moment, we will use the following definitions, which comply with the suggestions of Ref. [3]:

$$\tilde{S}_X(t) \equiv S_X(t) + \langle \beta^2 \partial_\beta E_x^*(\lambda_t) \rangle(t), \tag{52}$$

$$\tilde{Q}(t) \equiv \beta^{-1} \sum_{x,e} [p_{xe}(t) \ln \bar{p}_{e|x}(\lambda_t) - p_{xe}(0) \ln \bar{p}_{e|x}(\lambda_0)]$$
$$+ \langle \beta \partial_\beta E_x^*(\lambda_t) \rangle(t) - \langle \beta \partial_\beta E_x^*(\lambda_0) \rangle(0). \tag{53}$$

The latter can be also rewritten as

$$\tilde{Q}(t) = -W(t) + \langle \partial_\beta \beta E_x^*(\lambda_t) \rangle(t) - \langle \partial_\beta \beta E_x^*(\lambda_0) \rangle(0), \tag{54}$$

if we use the generally accepted definition for work $W(t) = \langle E_{xe}(\lambda_t) \rangle(t) - \langle E_{xe}(\lambda_0) \rangle(0)$. Assuming the first law of thermodynamics to be valid in the strong coupling case, this then implies a definition for internal energy:

$$\tilde{U}_X(t) = \langle \partial_\beta \beta E_x^*(\lambda_t) \rangle(t) = \langle E_x^* + \beta \partial_\beta E_x^* \rangle(t). \tag{55}$$

Introducing the nonequilibrium free energy

$$\tilde{F}_X(t) = \tilde{U}_X(t) - \frac{1}{\beta} \tilde{S}_X(t) = \langle E_x^*(\lambda_t) \rangle(t) - \frac{1}{\beta} S_X(t), \tag{56}$$

we can alternatively write Eq. (51) as

$$\tilde{\Sigma}(t) = \beta[W(t) - \Delta \tilde{F}_X(t)] \geqslant 0. \tag{57}$$

The definitions above of $\tilde{\Sigma}$, $\tilde{S}_X$, $\tilde{Q}$, $\tilde{U}_X$, and $\tilde{F}_X$ seem to provide a satisfactory extension of thermodynamics to the strong-coupling case and they coincide with the definitions used by Seifert, who further motivates them by arguments of equilibrium statistical mechanics [3]. In addition to Ref. [3], we have seen that the framework can be even extended by allowing for a time-dependence in the coupling $V_{xe}(t)$, too.

Nevertheless, the approach above should be taken with care because it is ambiguous [4], and is not formulated at the level of instantaneous rates implying that the positivity of entropy production Eq. (51) crucially relies on the choice of initial state Eq. (45).

B. Reduced thermodynamics description in $X \otimes Y$

We now clarify this situation by returning to our previous results in Sec. II C. where we assumed that the environment is made of two parts, $E = Y \otimes R$. The first part Y is strongly coupled to the system X and is explicitly described. The second part R is an ideal weakly coupled and Markovian thermal reservoir. Under these assumptions the ME Eq. (2) and the full Hamiltonian dynamics give rise to the same description in the reduced space $X \otimes Y$. This implies, e.g., that the work computed within the ME framework [see Eq. (5)] coincides with the work computed using the exact Hamiltonian dynamics as in Sec. III A.

In the limit of TSS, Y instantaneously equilibrates with respect to a given microstate of X. Thus, the initial requirement Eq. (45) is not only fulfilled initially but at any time t. This implicitly means that for any fixed value of λ_t, the global equilibrium steady state reads

$$p_{xyr} = \frac{e^{-\beta E_{xy}(\lambda_t)}}{\mathcal{Z}_{XY}(\lambda_t)} \frac{e^{-\beta E_r}}{\mathcal{Z}_R}, \qquad (58)$$

where E_r is the bare Hamiltonian of R. As a result, the HMF introduced in Eq. (46) *coincides* with

$$E_x^* = E_x - \frac{1}{\beta} \ln \langle e^{-\beta V_{xy}} \rangle_Y^{\text{eq}} = F_{Y|x} - \mathcal{F}_Y, \qquad (59)$$

which can be regarded as the HMF of $X \otimes Y$ only. Using the last equation together with Eqs. (37) and (38), it is not hard to deduce the following two relations:

$$\tilde{U}_X(t) = U_{XY}(t) - \mathcal{U}_Y, \qquad (60)$$

$$\tilde{S}_X(t) = S_{XY}(t) - \mathcal{S}_Y. \qquad (61)$$

Thus, apart from a time-independent additive constant the definitions for internal energy and system entropy *within TSS coincide* with the definitions based on the HMF. Furthermore, since both approaches agree on the definition of work, we can show for the heat flow that

$$\dot{\tilde{Q}}(t) = d_t \tilde{U}_X(t) - \dot{W}(t) = d_t U_{XY}(t) - \dot{W}(t) = \dot{Q}(t). \quad (62)$$

Thus, within TSS we agree on this definition too and are able to derive the first law at the level of instantaneous rates. Likewise, we can also prove the positivity of the entropy production *rate*

by noting that

$$\dot{\tilde{\Sigma}}(t) = d_t \tilde{S}_X(t) - \beta \dot{\tilde{Q}}(t) = d_t S_{XY}(t) - \beta \dot{Q}(t) = \dot{\Sigma}(t) \geqslant 0. \qquad (63)$$

As a preliminary summary, we have thus shown that within TSS, the framework introduced in Ref. [3] is thermodynamically consistent and can be greatly extended. Furthermore, no ambiguity is left within our approach, which allows us to refute the criticism raised in Ref. [4] for our setup.

It is interesting to ask what happens *away from TSS* when Y does not instantaneously conditionally equilibrate and $p_{y|x}$ is thus dynamically evolving. It is then possible to show that the framework of Sec. III A does *not* coincide with the original thermodynamic description of Sec. II anymore. For instance, we prove in Appendix B that the difference in entropy production can be expressed as

$$\tilde{\Sigma}(t) - \Sigma(t) = \beta(\Delta F_{XY} - \Delta \tilde{F}_X)$$
$$= D[p_{xy}(t) \| p_x(t)\bar{p}_{y|x}(\lambda_t)] \geqslant 0. \qquad (64)$$

Thus, $\tilde{\Sigma}(t)$ overestimates $\Sigma(t) = \int_0^t ds\, \dot{\Sigma}(s)$ by the relative entropy between the true state of $X \otimes Y$ and an idealized state of the form of Eq. (45). Also, the rate of change of $\tilde{\Sigma}(t)$ can be negative. This and other features are explicitly demonstrated in the next section with the help of an example where the ME description in $X \otimes Y$ exactly coincides with the reduced Hamiltonian dynamics.

To conclude, when it is possible to separate out the strongly coupled and non-Markovian degrees of freedom Y from the environment E, then the following hierarchy of inequalities holds,

$$\tilde{\Sigma}(t) \geqslant \Sigma(t) \geqslant \Sigma^{(1)}(t) \geqslant 0. \qquad (65)$$

The equality $\tilde{\Sigma}(t) = \Sigma(t) = \Sigma^{(1)}(t)$ holds in the limit of TSS. Each of the entropy production in Eq. (65) corresponds to a different layer of the description. $\Sigma^{(1)}(t)$ assumes Y to be conditionally (versus X) equilibrated and, of course, implicitly R to be equilibrated. $\Sigma(t)$ assumes only an ideal reservoir R and $\tilde{\Sigma}(t)$ is an exact result which can be applied to any Hamiltonian dynamics Eq. (44) as long as the initial condition is of the form of Eq. (45).

IV. DISCREPANCY IN THE NONMARKOVIAN REGIME

Within TSS, i.e., whenever the environment behaves Markovian by instantaneously adapting to the microstate of the system X, we have proven the equivalence of the coarse-grained thermodynamic framework from Sec. II C with the approach based on the HMF. In principle, both frameworks can be also applied beyond TSS and we will now provide a counterexample showing that the HMF-approach then no longer coincides with the standard framework of Sec. II.

We consider the example of driven Brownian motion, thereby demonstrating that our main results above do not only hold for dynamics on discrete states but also for continuous

variables. The global Hamiltonian with mass-weighted coordinates is specified by [21,22]

$$E_{xe}(t) = E_x(t) + V_{xe} + E_e, \tag{66}$$

$$E_x(t) = \tfrac{1}{2}\big[p_x^2 + \omega^2(t)x^2\big], \tag{67}$$

$$V_{xe} + E_e = \frac{1}{2}\sum_k\left[p_k^2 + v_k^2\left(x_k - \frac{c_k}{v_k^2}x\right)^2\right], \tag{68}$$

and we identify $\lambda_t = t$ in the following and use $\omega(t) = \omega_0 + g\sin(\omega_L t)$. We relaxed the notation meaning with $E_x(t)$ the energy associated to the microstate (x, p_x) and a microstate e of the bath is given by specifying (x_k, p_k) for all k. Furthermore, the spectral density (SD) of the bath is defined as and parametrized by

$$J(\omega) \equiv \frac{\pi}{2}\sum_k\frac{c_k^2}{v_k}\delta(\omega - v_k) = \frac{\lambda_0^2\gamma\omega}{\left(\omega^2 - \omega_1^2\right)^2 + \gamma^2\omega^2}. \tag{69}$$

Here, λ_0 controls the overall coupling strength between the system and the environment and γ changes the shape of the SD from a pronounced peak around ω_1 for small γ to a rather unstructured and flat SD for large γ.

The corresponding Langevin equation for this setup reads [21,22]

$$\ddot{x}(t) + \omega_0^2(t)x(t) + \int_0^t ds\,\Gamma(t - s)\dot{x}(s) = \xi(t), \tag{70}$$

with the friction kernel

$$\Gamma(t) \equiv \int_0^\infty d\omega\frac{2}{\pi\omega}J(\omega)\cos(\omega t) \tag{71}$$

and the noise $\xi(t)$, which obeys the statistics

$$\langle\xi(t)\rangle_E = 0, \quad \langle\xi(t)\xi(s)\rangle_E = \frac{1}{\beta}\Gamma(t - s). \tag{72}$$

We see that for an Ohmic SD $J(\omega) = \eta\omega$ (times a high-frequency cutoff as usual), we obtain $\Gamma(t) = 2\eta\delta(t)$, and this gives the standard Langevin equation with Gaussian white noise. Unfortunately, our SD Eq. (69) is not Ohmic unless we scale $\lambda_0 = \sqrt{\alpha_1\alpha_2\gamma}$, $\omega_1 = \sqrt{\alpha_2\gamma}$ and send $\gamma \to \infty$. This implies an Ohmic SD for sufficiently large α_2:

$$\lim_{\gamma\to\infty}J(\omega) = \alpha_1\frac{\alpha_2\omega}{\alpha_2^2 + \omega^2}. \tag{73}$$

Establishing a consistent thermodynamic framework for the general Langevin Eq. (70) cannot be done using standard tools from stochastic thermodynamics. One route, however, could be to take the definitions from Sec. III and to apply them here. Application of these definitions is facilitated by the fact that for a Brownian motion Hamiltonian the HMF *coincides* with the bare system Hamiltonian, i.e., $E_x^* = E_x$, which can be directly checked by evaluating the Gaussian integrals. Thus, the change in internal energy and system entropy read $\Delta\tilde{U}_X = \langle E_x(t)\rangle(t) - \langle E_x(0)\rangle(0)$ and $\Delta\tilde{S}_X = \langle -\ln p_x(t)\rangle(t) - \langle -\ln p_x(0)\rangle(0)$. That is to say, the HMF-approach uses for this examples the *standard* weak-coupling definitions *irrespective* of the spectral properties of the bath. Furthermore, since work can be computed using $W(t) = \int_0^t ds\langle d_t E_x(s)\rangle(s)$, we obtain $\tilde{Q}$ and $\tilde{\Sigma}$, too. However, to

access the dynamics of the system, we would have to simulate the non-Markovian Langevin Eq. (70), which is numerically demanding.

We therefore follow a different strategy and identify a subsystem $Y \subset E$, which transforms the non-Markovian system X to a Markovian system $X \otimes Y$. This is most conveniently done by identifying a collective degree of freedom in the environment defined via

$$\lambda_0 y \equiv \sum_k c_k x_k. \tag{74}$$

In this context, y is also known as a reaction coordinate. It has been shown to successfully model the dynamics of non-Markovian open quantum systems (see, e.g., Refs. [23–26]) and has been recently proposed as a method to establish a consistent thermodynamic framework beyond the Markovian and weak-coupling approximation [27,28].

We skip the details of the derivation, which can be looked up in the literature [23–28], and only state the main result. After the transformation, the Hamiltonian becomes

$$E_{xyr}(t) = E_x(t) + V_{xy} + E_y + V_{yr} + E_r, \tag{75}$$

$$V_{xy} + E_y = \frac{\lambda_0^2}{2\omega_1^2}x^2 - \lambda_0 xy + \frac{1}{2}\big(p_y^2 + \omega_1^2 y^2\big), \tag{76}$$

$$V_{yr} + E_r = \frac{1}{2}\sum_k\left[\tilde{p}_k^2 + \tilde{v}_k^2\left(\tilde{x}_k - \frac{\tilde{c}_k}{\tilde{v}_k^2}y\right)^2\right], \tag{77}$$

where the new SD of the "residual environment" R is defined and for the choice Eq. (69) given by

$$\tilde{J}(\omega) \equiv \frac{\pi}{2}\sum_k\frac{\tilde{c}_k^2}{\tilde{v}_k}\delta(\omega - \tilde{v}_k) = \gamma\omega. \tag{78}$$

This SD immediately yields the coupled set of *Markovian* Langevin equations

$$\begin{aligned}\ddot{x}(t) + \left[\omega_0^2(t) + \frac{\lambda_0^2}{\omega_1^2}\right]x(t) - \lambda_0 y(t) &= 0,\\ \ddot{y}(t) + \gamma\dot{y}(t) + \omega_1^2 y(t) - \lambda_0 x(t) &= \xi(t),\end{aligned} \tag{79}$$

with Gaussian white noise $\xi(t)$.

Following standard procedures [22], we can associate a Fokker-Planck equation for the probability distribution $P(t) = P(x, p_x, y, p_y; t)$ to the set of Langevin equations above. It reads

$$\partial_t P(t) = \big(-\nabla^T \cdot A \cdot \mathbf{x} + \tfrac{1}{2}\nabla^T \cdot B \cdot \nabla\big)P(t), \tag{80}$$

where we defined $\nabla \equiv (\partial_x, \partial_{p_x}, \partial_y, \partial_{p_y})^T$, $\mathbf{x} \equiv (x, p_x, y, p_y)^T$ and introduced the matrices

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\big[\omega_0^2(\lambda_t) + \lambda_0^2/\omega_1^2\big] & 0 & \lambda_0 & 0 \\ 0 & 0 & 0 & 1 \\ \lambda_0 & 0 & -\omega_1^2 & -\gamma \end{pmatrix} \tag{81}$$

and B whose only nonzero component is $B_{44} = 2\gamma/\beta$. We emphasize that Eq. (80) describes the exact dynamics in $X \otimes Y$. No approximation has been made in any of the steps above (apart from assuming an initially equilibrated reservoir state).

An advantage of this Fokker-Planck equation is that the dynamics of the first and second cumulants are closed. In fact, the equations of motion for the first cumulants $\langle z \rangle$ (with $z \in \{x, p_x, y, p_y\}$) couple only to themselves and the same is true for the second cumulants $C_{zz'} \equiv \langle zz' \rangle - \langle z \rangle \langle z' \rangle$. Thus, an initially Gaussian state will stay Gaussian for all times. Computing the time-evolution of the first two cumulants based on an initial condition of the form of Eq. (45) can then be easily done numerically. Because standard stochastic thermodynamics applies to Eq. (79) or (80), we have direct access to averaged thermodynamic quantities for $X \otimes Y$ introduced in Sec. II, also see Ref. [22]. Furthermore, because we have the exact dynamics in $X \otimes Y$, we also get the exact reduced dynamics of X by tracing over Y, consequently giving direct access to the time evolution of $\tilde{U}_X$, $\tilde{S}_X$, $\tilde{Q}$, and $\tilde{\Sigma}$. Thus, our Markovian embedding strategy has allowed us to circumvent the difficulty to simulate the non-Markovian Langevin Eq. (70).

Results of the simulation are shown in Fig. 2. We vary the SD from a strongly non-Markovian situation (shown on the left) to a Markovian but strong-coupling situation (on the right) by changing γ and α_1. For each γ, we compare the integrated heat flows Q and $\tilde{Q}$ (upper panel) and the integrated entropy production Σ and $\tilde{\Sigma}$ (lower panel). The following main features are observable: for large γ the assumption of TSS is justified and quantities defined in Sec. II C and III agree perfectly. In fact, γ is directly linked to the rate of relaxation of the reaction coordinate (y, p_y) but does not directly couple to the system degrees of freedom (x, p_x). Thus, a large γ corresponds to the limit of TSS as introduced in Eq. (25). Away from that limit, however, we observe that $\tilde{Q}$ differs *significantly* from Q and the same observation is true for the different definitions of entropy production, too. Also, although Eq. (51) is always obeyed, the rate of $\tilde{\Sigma}$ can become negative. Furthermore, we can also confirm the validity of Eq. (64), $\tilde{\Sigma}(t) \geqslant \Sigma(t)$.

V. SUMMARY

We clarified important questions in the framework of strong-coupling thermodynamics. Our main achievements are the following:

(1) Justification of the HMF within TSS. Within the limit of TSS, the framework provided in Ref. [3] is thermodynamically consistent for arbitrary system states $p_x(t)$ and the HMF is a legitimate tool to investigate the thermodynamics of systems in strong contact with a *single* environment.

(2) No ambiguity. Any ambiguity is removed in our framework and the concerns put forward in Ref. [4] do not apply. The reason for this is that we start from a well-defined weak coupling framework. Especially and contrary to previous attempts, we do not use the first law to define heat but have an alternative and unambiguous definition for it.

(3) Extension of previous results. Thanks to the TSS, we were able to significantly extend previous results by formulating them at the level of instantaneous rates instead of integrated quantities and by allowing also for a time-dependence in the system-environment coupling.

(4) Difficulties in the non-Markovian regime. Away from TSS, the framework of Ref. [3] does not match the original thermodynamic picture though Eq. (49) is always obeyed.

Thus, we observe that in order to establish the original laws of thermodynamics in the non-Markovian regime (where the environment is also dynamically evolving), one is forced to fully take into account the (thermo)dynamics of X *and* Y. Any effective description at this stage will in general miss important pieces in the first or second law. This complies with the point of view put forward in Ref. [27,28].

Recently, two alternative approaches were put forward in Ref. [6] by starting from the isothermal-isobaric ensemble and by taking pressure and volume effects into account. These approaches correctly reproduce the macroscopic limit by introducing the notion of "thermodynamic volume" for a microscopic system. The "bare representation" in Ref. [6] shows that it is possible to retain the original weak coupling definitions of internal energy and entropy by shifting our attention to enthalpy and Gibbs free energy instead. If the isobaric PV-contribution is absent or blindly ignored, then the "partial molar representation" in Ref. [6] coincides with the approach in Sec. III.[2] These alternative approaches should be therefore also derivable within TSS, but we expect that outside the limit of TSS they will mismatch again.

Finally, we mention that the framework of Sec. III can be used in principle also beyond TSS, for instance, if it is impossible to find a splitting $E = Y \otimes R$ or if the dynamical simulation of the environment becomes unfeasible. It then nevertheless has to be treated with care and further consistency checks still need to be carried out such as, for instance, the implication of the correct thermodynamic laws in the limit of reversible transformations as investigated in Ref. [29].

APPENDIX A: RELATION BETWEEN THE ENTROPY PRODUCTION IN EQ. (49) and Ref. [20]

In Ref. [20], the entropy production of Hamiltonian dynamics Eq. (44) with an initial condition of the form $p_x(0) p_e^{\text{eq}}$ was defined as

$$\Sigma_{\text{NJP}}(t) \equiv D\big(p_{xe}(t) \| p_x(t) p_e^{\text{eq}}\big) \geqslant 0. \qquad (A1)$$

This result is very close in spirit to the entropy production Eq. (49) that we derived in this paper for the same Hamiltonian dynamics but with an initial condition of the form of Eq. (45). It measures the deviation of the true state from an idealized product state where the environment is always at equilibrium instead of conditionally at equilibrium.

The only meaningful comparison between the two expressions requires us to consider situations where the two classes

---

[2]To compare notation, we have without PV-terms that the Gibbs free energies in Ref. [6] are related to our free energies via $G_0^{\mathcal{E}} = \mathcal{F}_E$, $G_x^{\mathcal{E}} = F_{E|x} - E_x$ and the solvation Hamiltonian of mean force becomes $\phi(x) = E_x^* - E_x$. However, note that the PV-term in Ref. [6] is actually only negligible at weak coupling. Then, $E_x^* \approx E_x$ and all the different frameworks coincide with the weak coupling limit.
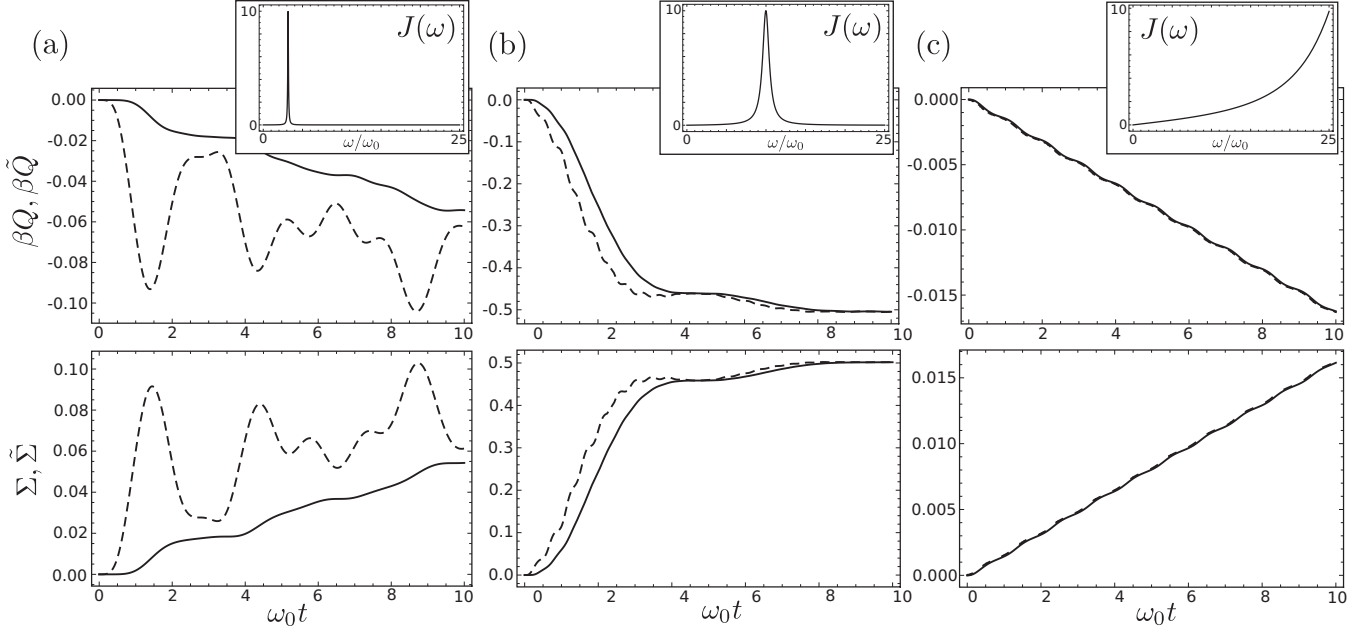
FIG. 2. Plot of the thermodynamics for a driven Brownian particle coupled to a non-Markovian environment. Each column (a), (b), (c) refers to a different form of the SD displayed in the upper right corner of each column. Below the SD, we show two different plots: the upper one displays $Q(t)$ (solid line) and $\tilde{Q}(t)$ (dashed line) as a function of dimensionless time $\omega_0 t$; the lower one displays $\Sigma(t)$ (solid line) and $\tilde{\Sigma}(t)$ (dashed line). Parameters for the driving are $\omega_L = 2\pi\omega_0$ and $g = 0.1\omega_0$. The SD is parametrized as $\lambda_0 = \sqrt{\alpha_1\alpha_2\gamma}$ and $\omega_1 = \sqrt{\alpha_2\gamma}$ with $\alpha_2 = 100\omega_0$. The SDs differ by the choice of (γ, α_1), which we chose as $(0.1\omega_0, \omega_0^2)$ in (a), $(\omega_0, 10^2\omega_0^2)$ in (b), and $(10\omega_0, 250^2\omega_0^2)$ in (c). The system was prepared using Eq. (45) and had initial mean values $\langle x \rangle(0) = (\sqrt{\beta}\omega_0)^{-1}$, $\langle p_x \rangle(0) = 0$ and covariances $C_{xx}(0) = (\beta\omega_0^2)^{-1}$, $C_{p_x p_x}(0) = \beta^{-1}$, and $C_{xp_x}(0) = 0$. Finally, we set $\omega_0 = 1$ and $\beta = 1$.

of initial conditions coincide, namely when $V_{xe}(\lambda_0) = 0$ and the interaction is only turned on afterwards. In this case, we find that

$$\tilde{\Sigma}(t) - \tilde{\Sigma}_{\text{NJP}}(t) = \sum_{x,e} p_{xe}(t) \ln \frac{p_e^{\text{eq}}}{\bar{p}_{e|x}(\lambda_t)}$$
$$= \beta \sum_{x,e} p_{xe}(t) \left[V_{xe}(\lambda_t) - \left(-\frac{1}{\beta} \ln \langle e^{-\beta V_{xe}(\lambda_t)} \rangle_E^{\text{eq}} \right) \right]. \tag{A2}$$

This relation can be rewritten as a difference between the nonequilibrium free energy Eq. (56) and the nonequilibrium free energy corresponding to the scheme of Ref. [20]:

$$F_{\text{NJP}}(t) \equiv \langle E_x(\lambda_t) + V_{xe}(\lambda_t) \rangle(t) - \beta^{-1} S_X(t). \tag{A3}$$

Explicitly,

$$\tilde{\Sigma}(t) - \tilde{\Sigma}_{\text{NJP}}(t) = \tilde{F}_X(t) - F_{\text{NJP}}(t). \tag{A4}$$

In general, there is no bound for this difference.

APPENDIX B: PROOF OF EQ. (64)

From Eq. (14), we deduce that $\Sigma(t) = \beta[W(t) - \Delta F_{XY}]$. Thus, with Eq. (57), we immediately get the first line of Eq. (64),

$$\tilde{\Sigma}(t) - \Sigma(t) = \beta(\Delta F_{XY} - \Delta \tilde{F}_X). \tag{B1}$$

To prove the second, we look at $F_{XY}(t) - \tilde{F}_X(t)$ and $F_{XY}(0) - \tilde{F}_X(0)$ in detail. Since the formalism using the HMF in Sec. III assumes that the environment starts in a conditionally

equilibrated state $p_{xy}(0) = p_x(0)\bar{p}_{y|x}(\lambda_0)$, Eqs. (60) and (61) are valid at $t = 0$. Straightforward algebra then gives

$$F_{XY}(0) - \tilde{F}_X(0) = \mathcal{F}_Y. \tag{B2}$$

At later times, using the definition Eqs. (52) and (55), we find that

$$F_{XY}(t) - \tilde{F}_X(t) = \langle E_{xy}(\lambda_t) \rangle - T S_{XY}(t)$$
$$- \langle E_x^*(\lambda_t) \rangle + T S_X(t). \tag{B3}$$

Next, from Eq. (59) together with Eq. (26), we obtain

$$\langle E_x^*(\lambda_t) \rangle = \langle E_{xy}(\lambda_t) \rangle + \beta^{-1} \langle \ln \bar{p}_{y|x}(\lambda_t) \rangle - \mathcal{F}_Y. \tag{B4}$$

Thus, we have explicitly

$$F_{XY}(t) - \tilde{F}_X(t) = \mathcal{F}_Y + \beta^{-1} \sum_{x,y} p_{xy}(t) \ln \frac{p_{y|x}(t)}{\bar{p}_{y|x}(\lambda_t)} \tag{B5}$$

and consequently,

$$\beta(\Delta F_{XY} - \Delta \tilde{F}_X) = \sum_{x,y} p_{xy}(t) \ln \frac{p_{y|x}(t)}{\bar{p}_{y|x}(\lambda_t)}, \tag{B6}$$

which proves the second line of Eq. (64).

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

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

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

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

[5] T. G. Philbin and J. Anders, Thermal energies of classical and quantum damped oscillators coupled to reservoirs, J. Phys. A: Math. Theor. **49**, 215303 (2016).

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

[7] H. J. D. Miller and J. Anders, Entropy production and time-asymmetry in the presence of strong interactions, arXiv:1703.03764 (2017).

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

[9] M. Campisi, P. Talkner, and P. Hänggi, Thermodynamics and fluctuation theorems for a strongly coupled open quantum system: An exactly solvable case, J. Phys. A.: Math Theor. **42**, 392002 (2009).

[10] S. Hilt, S. Shabbir, J. Anders, and E. Lutz, Landauer's principle in the quantum regime, Phys. Rev. E **83**, 030102 (2011).

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

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

[13] U. Seifert, Stochastic thermodynamics of single enzymes and molecular motors, Eur. Phys. J. E **34**, 26 (2011).

[14] S. Bo and A. Celani, Entropy production in stochastic systems with fast and slow time-scales, J. Stat. Phys. **154**, 1325 (2014).

[15] J. Schnakenberg, Network theory of microscopic and macroscopic behavior of master equation systems, Rev. Mod. Phys. **48**, 571 (1976).

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

[17] C. Van den Broeck and M. Esposito, Ensemble and trajectory thermodynamics: A brief introduction, Physica (Amsterdam) A **418**, 6 (2015).

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

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

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

[21] U. Weiss, *Quantum Dissipative Systems*, 3rd ed. (World Scientific, Singapore, 2008).

[22] K. Sekimoto, *Stochastic Energetics*, Lecture Notes in Physics, Vol. 799, 1st ed. (Springer-Verlag, Berlin, Heidelberg, 2010).

[23] J. Prior, A. W. Chin, S. F. Huelga, and M. B. Plenio, Efficient Simulation of Strong System-Environment Interactions, Phys. Rev. Lett. **105**, 050404 (2010).

[24] R. Martinazzo, B. Vacchini, K. H. Hughes, and I. Burghardt, Communication: Universal Markovian reduction of Brownian particle dynamics, J. Chem. Phys. **134**, 011101 (2011).

[25] M. P. Woods, R. Groux, A. W. Chin, S. F. Huelga, and M. B. Plenio, Mappings of open quantum systems onto chain representations and Markovian embeddings, J. Math. Phys. **55**, 032101 (2014).

[26] J. Iles-Smith, N. Lambert, and A. Nazir, Environmental dynamics, correlations and the emergence of noncanonical equilibrium states in open quantum systems, Phys. Rev. A **90**, 032114 (2014).

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

[28] D. Newman, F. Mintert, and A. Nazir, Performance of a quantum heat engine at strong reservoir coupling, Phys. Rev. E **95**, 032139 (2017).

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