Non-Markovianity and negative entropy production rates

Philipp Strasberg and Massimiliano Esposito

Physics and Materials Science Research Unit, University of Luxembourg, L-1511 Luxembourg, Luxembourg

Entropy production plays a fundamental role in nonequilibrium thermodynamics to quantify the irreversibility of open systems. Its positivity can be ensured for a wide class of setups, but the entropy production *rate* can become negative sometimes. This is often taken as an indicator of non-Markovianity. We make this link precise by showing under which conditions a negative entropy production rate implies non-Markovianity and when it does not. For a system coupled to a single heat bath, this can be established within a unified language for two setups: (i) the dynamics resulting from a coarse-grained description of a Markovian master equation and (ii) the classical Hamiltonian dynamics of a system coupled to a bath. The quantum version of the latter result is shown not to hold despite the fact that the integrated thermodynamic description is formally equivalent to the classical case. The instantaneous fixed point of a non-Markovian dynamics plays an important role in our study. Our key contribution is to provide a consistent theoretical framework to study the finite-time thermodynamics of a large class of dynamics with a precise link to its non-Markovianity.

I. INTRODUCTION

The theory of stochastic processes provides a powerful tool to describe the dynamics of open systems. Physically, the noise to which these systems are subjected results from the fact that the system is coupled to an environment composed of many other degrees of freedom about which we have only limited information and control. This coarse-grained description of the system, as opposed to the microscopic description involving the composite system *and* environment, is particularly appealing and tractable, when the Markovian approximation is applied. Therefore, Markovian stochastic dynamics are nowadays very commonly used to describe small open systems ranging from biochemistry (e.g., enzymes, molecular motors) to quantum systems (e.g., single atoms or molecules) [1–4]. Due to their outstanding importance for many branches of science, an entire branch of mathematics is also devoted to their study [5].

A common feature of all Markovian processes is their *contractivity*, i.e., the volume of accessible states shrinks monotonically during the evolution. This statement can be made mathematically precise by considering two arbitrary preparations $p_\alpha(0)$ and $q_\alpha(0)$, describing different probabilities to find the system in state α at the initial time $t = 0$. Their distance, as measured by the relative entropy $D[p_\alpha \| q_\alpha] \equiv \sum_\alpha p_\alpha \ln \frac{p_\alpha}{q_\alpha}$, monotonically decreases over time t, i.e., for all $t \geqslant 0$:

$$\frac{\partial}{\partial t} D[p_\alpha(t) \| q_\alpha(t)] \leqslant 0. \qquad (1)$$

In other words, the ability to distinguish between any pair of initial states monotonically shrinks in time due to a continuous loss of information from the system to the environment. We note that also other distance quantifiers than the relative entropy fulfill Eq. (1) and an analog of Eq. (1) also holds in the quantum regime where its violations have been proposed as an indicator of non-Markovianity [6–8].

The contractivity property (1) of Markov processes gets another interesting physical interpretation in quantum and stochastic thermodynamics. In these fields, a nonequilibrium thermodynamics is systematically built on top of Markovian dynamics typically described by (quantum) master or Fokker-Planck equations [9–16]. In addition to being Markovian, the rates entering the dynamics must also satisfy local detailed balance. For a system coupled to a single heat bath, this ensures that the Gibbs state of the system is a null eigenvector of the generator of the dynamics at all times t. For autonomous dynamics, this implies that the fixed point of the dynamics is an equilibrium Gibbs state. For nonautonomous (also called *driven*) dynamics, i.e., when some parameters are changed in time according to a prescribed protocol λ_t, the system in general does not reach a steady state, but the Gibbs state remains a null eigenvector of the generator of the dynamics at all times t. We call this an *instantaneous* fixed point of the dynamics in the following. If we denote the Gibbs state of the system by $e^{-\beta E_\alpha(\lambda_t)}/\mathcal{Z}(\lambda_t)$ with the energy $E_\alpha(\lambda_t)$ of state α and the equilibrium partition function $\mathcal{Z}(\lambda_t) = \sum_\alpha e^{-\beta E_\alpha(\lambda_t)}$, the second law of thermodynamics for a driven system in contact with a single heat bath at inverse temperature β can be expressed as

$$\dot{\Sigma}(t) = -\frac{\partial}{\partial t}\bigg|_{\lambda_t} D\left[p_\alpha(t) \left\| \frac{e^{-\beta E_\alpha(\lambda_t)}}{\mathcal{Z}(\lambda_t)} \right. \right] \geqslant 0. \qquad (2)$$

Here, the derivative is evaluated at fixed λ_t, i.e., $E_\alpha(\lambda_t)$ and $\mathcal{Z}(\lambda_t)$ are treated as constants, which only depend parametrically on time. The quantity $\dot{\Sigma}(t)$ is the entropy production rate. Its positivity follows from the fact that the dynamics is Markovian *and* that the Gibbs state is an instantaneous fixed point of the dynamical generator at all times. Within the conventional weak coupling and Markovian framework

[9–16], the entropy production rate can be rewritten as $\dot{\Sigma}(t) = \beta[\dot{W}(t) - d_t F(t)] \geqslant 0$, where $\dot{W}$ is the rate of work done on the system and $d_t F(t)$ denotes the change in nonequilibrium free energy (see Sec. III A for microscopic definitions of these quantities). The intimate connection between relative entropy and the second law was noticed some time ago in Ref. [17] for undriven systems. In the undriven case, the precise form of Eq. (2) seems to appear first in Ref. [18] for quantum systems and it is discussed as a Lyapunov function in Ref. [3] for classical systems. The generalization to driven systems was given in Ref. [19] and a similar form of Eq. (2) also holds for a system in contact with multiple heat baths [20] (see also Ref. [21] for a recent approach where Eq. (2) plays a decisive role). In this paper, we will only focus on a single heat bath.

While the Markovian assumption is widely used due to the enormous simplifications it enables, it is not always justified. Especially in stochastic thermodynamics an implicit but crucial assumption entering the Markovian description is that the degrees of freedom of the environment are always locally equilibrated with a well-defined associated temperature. This is in general only valid in the limit of timescale separation where the environmental degrees of freedom can be adiabatically eliminated [22]. There is currently no consensus about the correct thermodynamic description of a system when the local equilibrium assumption for the environment is not met, i.e., when the system dynamics is non-Markovian. Especially, while different interesting results were obtained in Refs. [23–28] by *starting* from a non-Markovian description of the system, the *emergence* of non-Markovianity and its link to an underlying Markovian description of the microscopic degrees of freedom (system *and* bath) was not yet established.

The first main contribution of this paper is to provide a systematic framework for that situation able to investigate the influence of an environment, which is not locally equilibrated. While there has been recently great progress in the *integrated* thermodynamic description of such systems [29–32], the instantaneous thermodynamic properties at the rate level were only studied in Ref. [32]. We will here see that a remarkably similar framework to the conventional one above arises with the main difference that the entropy production rate $\dot{\Sigma}(t)$ can be negative sometimes. We then precisely link the occurrence of $\dot{\Sigma}(t) < 0$ to underlying dynamical properties of the environment, thereby connecting the abstract mathematical property of (non-)Markovianity to an important physical observable.

Our second main contribution is to establish a quantum counterpart for the classical strong coupling scenario studied by Seifert [29]. We find that the integrated thermodynamic description is very similar, but the instantaneous rate level description is *not*. This hinders us to connect the occurrence of negative entropy production rates to the non-Markovianity of the system evolution. We also provide an explicit example to show that recent claims in the literature about non-Markovianity, negative entropy production rates, and steady states of dynamical maps do not hold.

This paper covers a wide range of applications from (i) rate master equations over (ii) classical Hamiltonian dynamics to (iii) quantum systems. We will keep this order in the
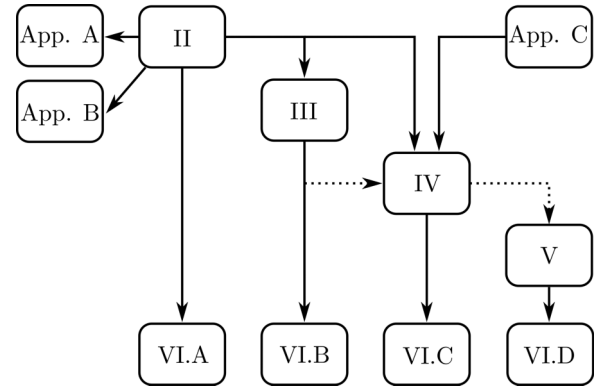

FIG. 1. "Road map" of the paper with solid (dotted) arrows indicating strong (weak) dependencies.

narrative because it demonstrates beautifully the similarities and discrepancies of the different levels of description. We will start with a purely mathematical description of classical, non-Markovian systems, which arise from an arbitrary coarse graining of an underlying Markovian network. While Sec. II A reviews known results, Sec. II B establishes theorems (Appendices A and B give additional technical details). Section III can then be seen as a direct physical application of the previous section to the coarse-grained dynamics of a Markovian network obeying local detailed balance. In Sec. IV we change the perspective and consider classical Hamiltonian system-bath dynamics, but with the help of Appendix C we will see that we obtain identical results to Sec. III. In our last general Sec. V we consider quantum systems. To illustrate the general theory, each subsection of Sec. VI is used to illustrate a particular feature of one of the previous sections. This road map of the paper is shown in Fig. 1 and we wish to emphasize that it is also possible to read some sections independently. The paper closes by summarizing our results together with the state of the art of the field in Sec. VII A and by discussing alternative approaches and open questions in Sec. VII B. We also provide an example to demonstrate that non-Markovian effects can speed up the erasure of a single bit of information, thereby showing that the field of non-Markovian finite-time thermodynamics provides a promising research direction for the future.

The following abbreviations are used throughout the text: EP (entropy production), IFP (instantaneous fixed point), ME (master equation), TM (transition matrix), and TSS (timescale separation).

II. MATHEMATICAL PRELIMINARIES

A. Coarse-grained Markov chains

In this section we establish notation and review some known results about Markov processes under coarse graining. We will start with the description of a discrete, time-homogeneous Markov chain for simplicity, but soon we will move to the physically more relevant case of an arbitrary continuous-time Markov process described by a ME. Finally, we also introduce the concept of lumpability [5].

Discrete, homogeneous Markov chains. We consider a Markov process on a discrete space $\mathcal{X}$ with N states $x \in \mathcal{X}$
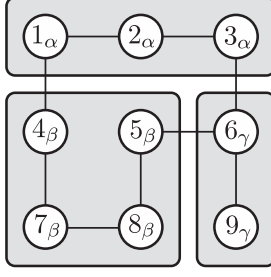
FIG. 2. Lumping (coarse graining) of a discrete Markov chain with microstate space $\mathcal{X} = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ into three mesostates according to the partition $\chi = \{\chi_\alpha, \chi_\beta, \chi_\gamma\}$ with $\chi_\alpha = \{1, 2, 3\}$, $\chi_\beta = \{4, 5, 7, 8\}$, and $\chi_\gamma = \{6, 9\}$ (gray areas). Possible transitions for which $T_\tau(x|y) \neq 0$ are depicted by a solid line connecting states x and y.

with a fixed TM $T_\tau(x|x')$, which propagates the state of the system such that

$$p_x(n\tau + \tau) = \sum_y T_\tau(x|y) p_y(n\tau) \quad (n \in \mathbb{N}), \quad (3)$$

or in vector notation $\mathbf{p}(n\tau + \tau) = T_\tau \mathbf{p}(n\tau)$. Here, $p_x(n\tau)$ is the probability to find the system in the state x at time $n\tau$, where $\tau > 0$ is an arbitrary but fixed time step (here and in what follows we will set the initial time to $t_0 \equiv 0$). Probability theory demands that $\sum_x p_x(n\tau) = 1$, $p_x(n\tau) \geqslant 0$ for all x, $\sum_x T_\tau(x|y) = 1$, and $T_\tau(x|y) \geqslant 0$ for all x, y. The steady state of the Markov chain is denoted by π_x and it is defined via the equation $\boldsymbol{\pi} = T_\tau \boldsymbol{\pi}$. In this section, we exclude the case of multiple steady states for definiteness, although large parts of the resulting theory can be applied to multiple steady states as well.[1]

Next, we consider a partition $\chi = \{\chi_1, \dots, \chi_M\}$ $(1 < M < N)$ of the state space such that

$$\bigcup_{\alpha=1}^M \chi_\alpha = \mathcal{X}, \quad \chi_\alpha \cap \chi_\beta = \emptyset \quad \text{for } \alpha \neq \beta. \quad (4)$$

In the physics literature this is known as a coarse-graining procedure where different "microstates" x are collected together into a "mesostate" α, whereas in the mathematical literature this procedure is usually called lumping. In the following, we will use both terminologies interchangeably and we denote a microstate x belonging to the mesostate α by x_α, i.e., $x_\alpha \in \chi_\alpha$. The idea is illustrated in Fig. 2. We remark that tracing out the degrees of freedom of some irrelevant system (usually called the "bath") is a special form of coarse graining. We will encounter this situation, e.g., in Sec. IV.

Any partition χ defines a stochastic process on the set of mesostates by considering for a given initial distribution $p_x(0)$ the probabilities to visit a sequence of mesostates $\alpha, \beta, \gamma, \dots$

---

[1]The contractivity property of Markov chains, Eqs. (1) and (2), which plays an important role in the following, holds true irrespective of the number of steady states.

at times $0, \tau, 2\tau, \dots$ with joint probabilities

$$p(\beta, \tau; \alpha, 0) = \sum_{y_\beta, x_\alpha} T_\tau(y_\beta|x_\alpha) p_{x|\alpha}(0) p_\alpha(0),$$

$$p(\gamma, 2\tau; \beta, \tau; \alpha, 0)$$
$$= \sum_{z_\gamma, y_\beta, x_\alpha} T_\tau(z_\gamma|y_\beta) T_\tau(y_\beta|x_\alpha) p_{x|\alpha}(0) p_\alpha(0), \quad (5)$$

etc., where $p_\alpha(0) = \sum_{x_\alpha} p_{x_\alpha}(0)$ is the marginalized initial mesostate and $p_{x|\alpha}(0) = p_{x_\alpha}(0)/p_\alpha(0)$ is the initial microstate conditioned on a certain mesostate α. The so generated hierarchy of joint probabilities $p(\alpha_n, n\tau; \dots; \alpha_1, \tau; \alpha_0, 0)$ completely specifies the stochastic process at the mesolevel. It is called Markovian whenever the conditional probabilities

$$p(\alpha_n, n\tau|\alpha_{n-1}, n\tau - \tau; \dots; \alpha_0, 0)$$

$$\equiv \frac{p(\alpha_n, n\tau; \dots; \alpha_0, 0)}{p(\alpha_{n-1}, n\tau - \tau; \dots; \alpha_0, 0)} \quad (6)$$

satisfy the Markov property [3,5,7,8]

$$p(\alpha_n, n\tau|\alpha_{n-1}, n\tau - \tau; \dots; \alpha_0, 0)$$
$$= p(\alpha_n, n\tau|\alpha_{n-1}, n\tau - \tau). \quad (7)$$

In practice, this requires to check infinitely many conditions. But, as we will see below, to compute all quantities of thermodynamic interest, only the knowledge about the evolution of the *one*-time probabilities $p(\alpha_n, n\tau)$ is important for us.

To see how non-Markovianity affects the evolution of the one-time probabilities, we introduce the following matrices derived from the above joint probabilities:

$$G_{\tau,0}(\beta|\alpha) = \frac{p(\beta, \tau; \alpha, 0)}{p_\alpha(0)} = \sum_{y_\beta, x_\alpha} T_\tau(y_\beta|x_\alpha) p_{x|\alpha}(0),$$

$$\tilde{G}_{2\tau,\tau}(\gamma|\beta) = \frac{p(\gamma, 2\tau; \beta, \tau)}{p_\beta(\tau)} = \frac{\sum_\alpha p(\gamma, 2\tau; \beta, \tau; \alpha, 0)}{\sum_\alpha p(\beta, \tau; \alpha, 0)},$$

$$G_{2\tau,0}(\gamma|\alpha) = \frac{p(\gamma, 2\tau; \alpha, 0)}{p_\alpha(0)}$$

$$= \sum_{z_\gamma, x_\alpha} \sum_{\beta, y_\beta} T_\tau(z_\gamma|y_\beta) T_\tau(y_\beta|x_\alpha) p_{x|\alpha}(0). \quad (8)$$

Formally, these matrices are well-defined conditional probabilities because they are positive and normalized. However, we have deliberately chosen a different notation for $\tilde{G}_{2\tau,\tau}$ because only $G_{\tau,0}$ and $G_{2\tau,0}$ can be interpreted as *transition* probabilities (or matrices) as they generate the correct time evolution for *any* initial mesostate $p_\alpha(0)$. The matrix $\tilde{G}_{2\tau,\tau}$ instead depends on the specific choice of $p_\alpha(0)$: if we start with a different initial mesostate $q_\alpha(0) \neq p_\alpha(0)$, we cannot use $\tilde{G}_{2\tau,\tau}$ to propagate $q_\beta(\tau) = \sum_\beta G_{\tau,0}(\beta|\alpha) q_\alpha(0)$ further in time. This becomes manifest by realizing that the so generated hierarchy of conditional probabilities does not in general obey the Chapman-Kolmogorov equation

$$G_{2\tau,0}(\gamma|\alpha) = \sum_\beta \tilde{G}_{2\tau,\tau}(\gamma|\beta) G_{\tau,0}(\beta|\alpha). \quad (9)$$

A way to avoid this undesired feature is to define the TM from time τ to 2τ via the inverse of $G_{\tau,0}$ (provided it exists) [7,8,33,34]:

$$G_{2\tau,\tau} \equiv G_{2\tau,0} G_{\tau,0}^{-1}. \tag{10}$$

The TM $G_{2\tau,\tau}$ does not depend on the initial mesostate, preserves the normalization of the state, and, by construction, it fulfills the Chapman-Kolmogorov equation $G_{2\tau,0} = G_{2\tau,\tau} G_{\tau,0}$. However, as the inverse of a positive matrix is not necessarily positive, $G_{2\tau,\tau}$ can have negative entries. This clearly indicates that $G_{2\tau,\tau}(\gamma|\beta)$ cannot be interpreted as a conditional probability and, hence, the process must be non-Markovian. Based on these insights, we introduce a weaker notion of Markovianity, which we coin 1-Markovianity. In the context of open quantum systems dynamics, this notion is often simply called Markovianity [7,8]:

Definition 1 (1-Markovianity). A stochastic process is said to be 1-Markovian, if the set of TMs $\{G_{n\tau,m\tau}|n \geqslant m \geqslant 0\}$ introduced above fulfill $G_{n\tau,m\tau}(\alpha|\beta) \geqslant 0$ for all $n \geqslant m \geqslant 0$ and all α, β.

It is important to realize that the notion of 1-Markovianity is weaker than the notion of Markovianity: if the coarse-grained process is Markovian, then it is also 1-Markovian and the TMs coincide with the conditional probabilities in Eq. (7). Furthermore, there exist processes which are 1-Markovian but not Markovian according to Eq. (7) (see, e.g., Ref. [7]).

Before we consider MEs, we introduce some further notation. We let

$$\mathcal{A}(0) \equiv \{p_x(0)|p_\alpha(0) \text{ arbitrary, } p_{x|\alpha}(0) \text{ fixed}\} \tag{11}$$

be the set of all *physically admissible initial states* with respect to a partition $\boldsymbol{\chi}$ (whose dependence is implicit in the notation). The reason to keep $p_{x|\alpha}(0)$ fixed is twofold: first, in an experiment one usually does not have detailed control over the microstates, and second, the TMs (8) for the lumped process depend on $p_{x|\alpha}(0)$, i.e., every choice of $p_{x|\alpha}(0)$ defines a different stochastic process at the mesolevel and should be treated separately. Which of the mesostates $p_\alpha(0)$ we can really prepare in an experiment is another interesting (but for us unimportant) question; sometimes this could be only a single state (e.g., the steady state π_α). Of particular importance for the applications later on will be the set

$$\mathcal{A}_\pi \equiv \{p_\alpha \pi_{x|\alpha}|p_\alpha \text{ arbitrary}\}, \tag{12}$$

where $\pi_{x|\alpha} = \pi_{x_\alpha}/\pi_\alpha$ is the conditional steady state. Experimentally, such a class of states can be prepared by holding the mesostate fixed while allowing the microstates to reach steady state. Finally, we define the set of time-evolved admissible initial states

$$\mathcal{A}(\tau) \equiv \{\mathbf{p}(\tau) = T_\tau \mathbf{p}(0)|\mathbf{p}(0) \in \mathcal{A}(0)\}. \tag{13}$$

Time-dependent MEs. For many physical applications it is indeed easier to derive a ME, which describes the continuous-time evolution of the system state, compared to deriving a TM

for a finite time step [3,4]. The ME reads as in general

$$\frac{\partial}{\partial t} p_x(t) = \sum_y W_{x,y}(\lambda_t) p_y(t) \tag{14}$$

or in vector notation $\partial_t \mathbf{p}(t) = W(\lambda_t) \mathbf{p}(t)$. The rate matrix $W(\lambda_t)$ fulfills $\sum_x W_{x,y}(\lambda_t) = 0$ and $W_{x,y}(\lambda_t) \geqslant 0$ for $x \neq y$ and it is now also allowed to be parametrically dependent on time through a prescribed parameter λ_t. This situation usually arises by subjecting the system to an external drive, e.g., a time-dependent electric or magnetic field. Furthermore, we assume that the rate matrix has one IFP, which fulfills $W(\lambda_t) \boldsymbol{\pi}(\lambda_t) = 0$. Clearly, the steady state will in general also parametrically depend on λ_t.

We can connect the ME description to the theory above by noting that the TM over any finite time interval $[t, t + \tau]$ is formally given by

$$T_{t,t+\tau} = \mathcal{T}_+ \exp \int_t^{t+\tau} W(\lambda_s) ds, \tag{15}$$

where $\mathcal{T}_+$ is the time-ordering operator. In particular, if we choose $\delta t = \tau/N$ small enough such that $\lambda_{t+\delta t} \approx \lambda_t$ (assuming that λ_t changes continuously in time), we can approximate the TM to any desired accuracy via

$$T_{t+\tau,t} \approx \prod_{i=0}^{N-1} T_{t+i\delta t+\delta t, t+i\delta t} \equiv \prod_{i=0}^{N-1} e^{W(\lambda_{t+i\delta t})\delta t}. \tag{16}$$

As a notational convention, whenever the system is undriven (i.e., $\dot{\lambda}_t = 0$ for all t), we will simply drop the dependence on λ_t in the notation.

We now fix an arbitrary partition $\boldsymbol{\chi}$ as before. To describe the dynamics at the mesolevel, one can use several formally exact procedures, two of them we mention here. First, from Eq. (14) we get by direct coarse graining

$$\frac{\partial}{\partial t} p_\alpha(t) = \sum_\beta R_{\alpha,\beta}[\lambda_t, p_\alpha(0)] p_\beta(t),$$

$$R_{\alpha,\beta}[\lambda_t, p_\alpha(0)] \equiv \sum_{x_\alpha, y_\beta} W_{x_\alpha, y_\beta}(\lambda_t) p_{y|\beta}(t). \tag{17}$$

Here, the matrix $R[\lambda_t, p_\alpha(0)]$ still fulfills all properties of an ordinary rate matrix: $\sum_\alpha R_{\alpha,\beta}[\lambda_t, p_\alpha(0)] = 0$ and $R_{\alpha,\beta}[\lambda_t, p_\alpha(0)] \geqslant 0$ for $\alpha \neq \beta$. However, it explicitly depends on the initial mesostate $p_\alpha(0)$, which influences $p_{y|\beta}(t)$ for later times t. This is analogous to the problem mentioned below Eq. (8): the TMs computed with Eq. (17) at intermediate times depend on the initial state of the system. This reflects the non-Markovian character of the dynamics and makes it inconvenient for practical applications. Note that Eq. (17) still requires to solve for the full microdynamics and does not provide a closed reduced dynamical description.

A strategy to avoid this undesired feature follows the logic of Eq. (10) and only makes use of the well-defined transition probability [cf. Eq. (8)]

$$G_{t,0}(\alpha|\beta) \equiv \sum_{x_\alpha, y_\beta} T_{t,0}(x_\alpha|y_\beta) p_{y|\beta}(0). \tag{18}$$

Provided that its inverse exists,[2] it allows to define an effective ME independent of the initial mesostate [7,8,33,34]

$$\frac{\partial}{\partial t} p_\alpha(t) = \sum_\beta V_{\alpha,\beta}(\lambda_t, t) p_\beta(t), \tag{19}$$

$$V(\lambda_t, t) \equiv \lim_{\delta t \to 0} \frac{G_{t+\delta t, 0} G_{t,0}^{-1} - 1}{\delta t}, \tag{20}$$

but where the matrix $V(\lambda_t, t)$ now carries an additional time dependence, which does not come from the parameter λ_t. Notice that the construction (20) shares some similarity with the time convolutionless ME derived from the Nakajima-Zwanzig projection operator formalism, which is another formally exact ME independent of the initial mesostate [4,37–39]. The generator $V(\lambda_t, t)$ preserves normalization and yields to a set of TMs, which fulfill the Chapman-Kolmogorov equation, but it can have temporarily negative rates, i.e., $V_{\alpha,\beta}(\lambda_t, t) < 0$ for $\alpha \neq \beta$ is possible. This is a clear indicator that the dynamics is not 1-Markovian [40].

Finally, we note that there are also other MEs to describe the reduced state of the dynamics, e.g., the standard Nakajima-Zwanzig equation which is an integrodifferential equation [4,39]. This ME is free from the assumption that the inverse of Eq. (18) exists and therefore more general. On the other hand, we will see in Sec. II B that we will need the notion of an IFP of the dynamics, which is hard to define for an integrodifferential equation.

Lumpability. In this final part, we introduce the concept of lumpability from Sec. 6.3 in Ref. [5]. It will help us to further understand the conditions which ensure Markovianity at the mesolevel and it will be occasionally used in the following. In unison with Ref. [5] we first introduce the concept for discrete, time-homogeneous Markov chains before we consider MEs again. Furthermore, we emphasize that in the definition below, the notion of Markovianity refers to the usual property (7) and not only to the one-time probabilities. Another related weaker concept (known as "weak lumpability") is treated for the interested reader in Appendix A.

Definition 2 (Lumpability). A Markov chain with TM T_τ is lumpable with respect to a partition χ if for every initial distribution $p_x(0)$ the lumped process is a Markov chain with transition probabilities independent of $p_x(0)$.

It follows from the definition that a lumpable process for a given TM T_τ and partition χ is also a lumpable process for all larger times, i.e., for all $T_{n\tau} = (T_\tau)^n$ with $n > 1$ and the same partition χ. The following theorem will be useful for us:

Theorem 1. A necessary and sufficient condition for a Markov chain to be lumpable with respect to the partition χ is that

$$\mathcal{G}_\tau(\alpha|\beta) \equiv \sum_{x_\alpha} T_\tau(x_\alpha|y_\beta) = \sum_{x_\alpha} T_\tau(x_\alpha|y'_\beta) \tag{21}$$

---

[2]Finding a general answer to the question as to whether the inverse of a dynamical map exists, which allows one to construct a time-local ME, is nontrivial. Nevertheless, many open systems can be described by a time-local ME and this assumption seems to be less strict than one might initially guess. See Refs. [35,36] for further research on this topic.

holds for any $y_\beta \neq y'_\beta$. The lumped process then has the TM $\mathcal{G}_\tau$.

The details of the proof can be found in Ref. [5]. However, it is obvious that the so-defined set of TMs is independent of the initial state. In addition, one can readily check that they fulfill the Chapman-Kolmogorov equation, are normalized, and have positive entries.

The concept of lumpability can be straightforwardly extended to time-dependent MEs by demanding that a lumpable ME with respect to the partition χ has lumpable TMs $T_{t+\delta t, t}$ for any time t and every $\delta t > 0$. By expanding Eq. (21) in δt and by taking $\delta t \to 0$, we obtain the following corollary (see also Ref. [41]):

Corollary 1. A ME with possibly time-dependent rates is lumpable with respect to the partition χ if and only if

$$\mathcal{V}_{\alpha,\beta}(\lambda_t) \equiv \sum_{x_\alpha} W_{x_\alpha, y_\beta}(\lambda_t) = \sum_{x_\alpha} W_{x_\alpha, y'_\beta}(\lambda_t) \tag{22}$$

for any $y_\beta \neq y'_\beta$ and any t. The lumped process is then governed by the rate matrix $\mathcal{V}(\lambda_t)$.

Notice that the dynamical description of a lumpable ME is unambiguous because the generator $R[\lambda_t, p_\alpha(0)]$ from Eq. (17) and $V(\lambda_t, t)$ from Eq. (20) both coincide with $\mathcal{V}(\lambda_t)$ from the above corollary. For $R[\lambda_t, p_\alpha(0)]$ this follows from directly applying Eq. (22) to Eq. (17). For $V(\lambda_t, t)$, this follows from the fact that the propagator in Eq. (10) coincides for a Markovian process with the transition probabilities obtained from Eq. (7), which for a lumpable process are identical to the TMs introduced in Theorem 1. All generators are then identical and have the same well-defined rate matrix.

In the following, we will stop repeating that any concept at the coarse-grained level is always introduced "with respect to the partition χ."

B. Entropy production rates, non-Markovianity, and instantaneous fixed points

After having discussed how to describe the dynamics at the mesolevel, we now turn to its thermodynamics. This is still done in an abstract way without recourse to an underlying physical model. An important concept in our theory is the notion of an IFP, which we define as follows:

Definition 3 (Instantaneous fixed point). Let $V(\lambda_t, t)$ be the generator of the time-local ME (19). We say that $\tilde{\pi}(t)$ is an IFP of the dynamics if $V(\lambda_t, t)\tilde{\pi}(t) = 0$.

We notice that $\tilde{\pi}(t)$ does not need to be a well-defined probability distribution because $V(\lambda_t, t)$ can have negative rates. We also point out that the IFP at time t might not be reachable from any state in the class of initially admissible states and it is therefore a purely abstract concept. Hence, while $V(\lambda_t, t)\tilde{\pi}(t) = 0$ it need not be true that $R[\lambda_t, p_\alpha(0)]\tilde{\pi}(t) = 0$ for any $p_{x_\alpha}(0) \in \mathcal{A}(0)$. The IFP cannot be computed with the help of the effective rate matrix in Eq. (17). The IFP is only well defined for a time-local ME with a generator independent of the initial mesostate. In Appendix B, we will show that it also does not matter how we have derived the ME as long as it is time local, formally exact, and independent of the initial mesostate.

Now, we introduce the concept of EP rate in a formal way and establish a general theorem. Afterwards, we will

answer the question when does the IFP $\tilde{\boldsymbol{\pi}}(t)$ coincide with the marginalized IFP of the microdynamics,

$$\pi_\alpha(\lambda_t) = \sum_{x_\alpha} \pi_{x_\alpha}(\lambda_t). \qquad (23)$$

EP rate. We define the EP rate for the coarse-grained process by

$$\dot{\Sigma}(t) \equiv -\left.\frac{\partial}{\partial t}\right|_{\lambda_t} D[p_\alpha(t)\|\pi_\alpha(\lambda_t)]$$

$$= -\sum_\alpha \frac{\partial p_\alpha(t)}{\partial t}[\ln p_\alpha(t) - \ln \pi_\alpha(\lambda_t)], \qquad (24)$$

where $\pi_\alpha(\lambda_t)$ was defined in Eq. (23).[3] Notice that $\dot{\Sigma}(t)$ can be defined for any stochastic process and *a priori* it is not related to the physical EP rate known from nonequilibrium thermodynamics. However, for the systems considered in Secs. III and IV this will turn out to be the case. Having emphasized this point, we decided for simplicity to refrain from introducing a new terminology for $\dot{\Sigma}(t)$ in this section. Furthermore, we remark that the definition of $\dot{\Sigma}(t)$ is experimentally meaningful: it only requires to measure the mesostate $p_\alpha(t)$ and the knowledge of $\pi_\alpha(\lambda_t)$. The latter can be obtained by measuring the steady state of the system after holding λ_t fixed for a long time or by arguments of equilibrium statistical mechanics (see Secs. III and IV). Also, theoretically, Eq. (24) can be evaluated with any method that gives the exact evolution of the mesostates.

The following theorem shows how to connect negative EP rates to non-Markovianity. Application of this theorem to various physical situations will be the purpose of the next sections.

Theorem 2. If $\pi_\alpha(\lambda_t)$ is an IFP of the mesodynamics and if I denotes the time interval in which the mesodynamics are 1-Markovian, then $\dot{\Sigma}(t) \geqslant 0$ for all $t \in I$.

To prove this theorem, it is useful to recall the well-known lemma, which we have stated already in Eq. (1):

Lemma 1. For a 1-Markovian process, the relative entropy between any two probability distributions is continuously decreasing in time, i.e., for all t and any pair of initial distributions $p_\alpha(0)$ and $q_\alpha(0)$ Eq. (1) holds.

This lemma follows from the fact that, first, for every stochastic matrix M and any pair of distributions p_α and q_α one has that

$$D\left[\sum_\beta M_{\alpha,\beta}p_\beta \left\| \sum_\beta M_{\alpha,\beta}q_\beta \right.\right] \leqslant D[p_\alpha\|q_\alpha], \qquad (25)$$

and, second, for a 1-Markovian process the TM at any time t and for every time step δt is stochastic. We can now prove Theorem 2:

─────────

[3]We remark that it turns out to be important to use in our definition (24) the coarse-grained steady state $\pi_\alpha(\lambda_t)$ and not the actual IFP $\tilde{\pi}_\alpha(t)$ of the generator $V(\lambda_t, t)$. In the latter case, the so-defined EP rate has only a clear thermodynamic meaning in the Markovian limit, where it was previously identified with the nonadiabatic part of the EP rate [42,43].

Proof. By definition of the EP rate we have

$$\dot{\Sigma}(t)$$

$$= -\lim_{\delta t \to 0} \frac{D[G_{t+\delta t,t}\mathbf{p}_{\text{cg}}(t)\|\boldsymbol{\pi}_{\text{cg}}(\lambda_t)] - D[\mathbf{p}_{\text{cg}}(t)\|\boldsymbol{\pi}_{\text{cg}}(\lambda_t)]}{\delta t},$$

$$(26)$$

where $G_{t+\delta t,t}$ is the propagator obtained from the ME (19) [cf. also Eq. (10)], $\mathbf{p}_{\text{cg}}(t)$ denotes the vector of the coarse-grained state $p_\alpha(t)$ and likewise for $\boldsymbol{\pi}_{\text{cg}}(\lambda_t)$. Next, we use the assumption that $\boldsymbol{\pi}_{\text{cg}}(\lambda_t)$ is an IFP of the ME (19), i.e., we have

$$G_{t+\delta t,t}\boldsymbol{\pi}_{\text{cg}}(\lambda_t) \approx \boldsymbol{\pi}_{\text{cg}}(\lambda_t) \qquad (27)$$

and any possible discrepancy vanishes in the limit $\delta t \to 0$. Thus, we can rewrite Eq. (26) as

$$\dot{\Sigma}(t) = -\lim_{\delta t \to 0}\frac{1}{\delta t}\{D[G_{t+\delta t,t}\mathbf{p}_{\text{cg}}(t)\|G_{t+\delta t,t}\boldsymbol{\pi}_{\text{cg}}(\lambda_t)]$$

$$- D[\mathbf{p}_{\text{cg}}(t)\|\boldsymbol{\pi}_{\text{cg}}(\lambda_t)]\}. \qquad (28)$$

Now, if the dynamics is 1-Markovian (Definition 1), then $G_{t+\delta t,t}$ is a stochastic matrix and from Eq. (25) it follows that $\dot{\Sigma}(t) \geqslant 0$. ∎

Whereas the proof of Theorem 2 is straightforward, two things make it a nontrivial statement. First, we will show that the EP rate defined in Eq. (24) deserves its name because it can be linked to *physical* quantities with a *precise thermodynamic interpretation*. This will be done in Secs. III and IV. Second, the essential assumption that $\pi_\alpha(\lambda_t)$ is an IFP of the mesodynamics is nontrivial: it is *not* a consequence of a 1-Markovian time evolution and it can also happen for *non*-Markovian dynamics. The details of this crucial assumption will be worked out in the remainder of this section, but already at this point we emphasize that 1-Markovianity alone is *not* sufficient to guarantee that $\dot{\Sigma}(t) \geqslant 0$. The Venn diagram in Fig. 3 should help to understand the implications of Theorem 2 better.

IFP of the coarse-grained process. To answer the question when is $\tilde{\pi}_\alpha(\lambda_t) = \pi_\alpha(t)$, we start with the simple case and assume that the coarse-grained dynamics is lumpable. Hence, according to Corollary 1 there is a unique and well-defined rate matrix. We then get the following theorem:

Theorem 3. If the stochastic process is lumpable for some time interval I, then the IFP of the mesostates is given by the marginal IFP of $W(\lambda_t)$ for all $t \in I$.

Proof. We want to show that $\mathcal{V}(\lambda_t)\boldsymbol{\pi}(\lambda_t) = 0$. By using Corollary 1 in the first and third equality, we obtain

$$\sum_\beta \mathcal{V}_{\alpha,\beta}(\lambda_t)\pi_\beta(\lambda_t) = \sum_\beta \sum_{x_\alpha} W_{x_\alpha,y_\beta}(\lambda_t)\pi_\beta(\lambda_t)$$

$$= \sum_\beta \sum_{x_\alpha,y'_\beta} W_{x_\alpha,y_\beta}(\lambda_t)\pi_{y'_\beta}(\lambda_t)$$

$$= \sum_\beta \sum_{x_\alpha,y'_\beta} W_{x_\alpha,y'_\beta}(\lambda_t)\pi_{y'_\beta}(\lambda_t), \qquad (29)$$

which is zero since $\pi_{y'_\beta}(\lambda_t)$ is the IFP at the microlevel. ∎

Therefore, together with Theorem 2 we can infer that $\dot{\Sigma}(t) < 0$ unambiguously shows that the dynamics is not lumpable. However, lumpability required the coarse-grained
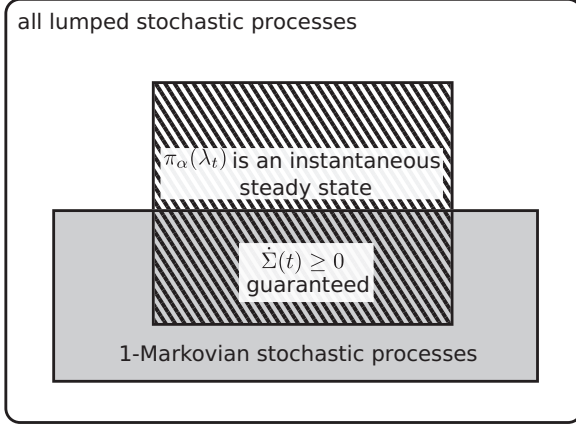
FIG. 3. A Venn diagram to understand the implications of Theorem 2. The largest outer box contains all possible lumped stochastic process. One subset of them is 1-Markovian (shaded in gray). For another subset the marginalized microlevel steady state $\pi_\alpha(\lambda_t)$ is an IFP of the dynamics (striped area). Where both sets overlap, $\dot{\Sigma}(t) \geqslant 0$ is guaranteed, i.e., whenever we observe $\dot{\Sigma}(t) < 0$ we cannot be simultaneously in the striped and in the shaded gray areas. Note that the Venn diagram shows the situation for a fixed driving protocol λ_t and time interval I. Depending on λ_t and I, the shaded gray and the striped areas can parametrically change in time.

process to fulfill the Markov property (7) for any initial condition, which is a rather strong property. We are therefore interested whether a negative EP rate reveals also insights about the weaker property of 1-Markovianity. For instance, for undriven processes we intuitively expect that, provided that we start at steady state, we always remain at steady state independently of the time dependence of the generator (20) or even the question whether the inverse of Eq. (18) exists. Then, negative values of the EP rate will always indicate non-Markovian dynamics for undriven system. Indeed, the following theorem holds:

Theorem 4. Consider an undriven stochastic process described by the ME (19), i.e., we assume $G_{t,0}^{-1}$ to exist for all admissible initial states $\mathcal{A}(0)$ and all times t. If the conditional microstates are initially equilibrated, $\mathcal{A}(0) \subset \mathcal{A}_\pi$ [Eq. (12)], then π_α is an IFP of the stochastic process at the mesolevel.

Proof. If $\mathcal{A}(0) \subset \mathcal{A}_\pi$, we can conclude that $\sum_\beta G_{t,0}(\alpha|\beta)\pi_\beta = \pi_\alpha$, i.e., if we start with the coarse-grained steady state we also remain in it for all times t. Since $G_{t,0}$ was assumed to be invertible,

$$\sum_\beta \left(G_{t,0}^{-1}\right)_{\alpha,\beta} \pi_\beta = \pi_\alpha. \tag{30}$$

Hence, by definition (20) we obtain the chain of equalities

$$\sum_{\alpha,\beta} \left(\lim_{\delta t \to 0} \frac{G_{t+\delta t,0} G_{t,0}^{-1} - 1}{\delta t} \right)_{\alpha,\beta} \pi_\beta$$

$$= \lim_{\delta t \to 0} \frac{1}{\delta t} \left[\sum_{\beta,\gamma} G_{t+\delta t,0}(\alpha|\gamma) \left(G_{t,0}^{-1}\right)_{\gamma,\beta} \pi_\beta - \pi_\alpha \right]$$

$$= \lim_{\delta t \to 0} \frac{1}{\delta t} \left[\sum_\gamma G_{t+\delta t,0}(\alpha|\gamma) \pi_\gamma - \pi_\alpha \right]$$

$$= \lim_{\delta t \to 0} \frac{1}{\delta t} [\pi_\alpha - \pi_\alpha] = 0. \tag{31}$$

■

We recognize a big difference in the characterization of the IFPs for driven and undriven processes. Without driving, the right set of initial states suffices already to show that the microlevel steady state induces the steady state at the mesolevel, even if the dynamics is non-Markovian. Thus, for this kind of dynamics $\dot{\Sigma}(t) < 0$ unambiguously signifies non-Markovianity. For driven systems instead, we needed the much stronger requirement of lumpability, i.e., Markovianity of the lumped process with TMs independent of the initial microstate. However, at least formally it is possible to establish the following additional theorem:

Theorem 5. Consider a driven stochastic process described by the ME (19), i.e., we assume $G_{t,0}^{-1}$ to exist for all initial states and all times t. We denote by I the time interval in which either

(1) all conditional microstates in the set of time-evolved states are at steady state $\mathcal{A}(t) \subset \mathcal{A}_\pi(\lambda_t)$, or

(2) the IFP of the microdynamics is an admissible time-evolved state $\pi_x(\lambda_t) \in \mathcal{A}(t)$.

Then, $\pi_\alpha(\lambda_t)$ is an IFP of the lumped process for all $t \in I$.

Proof. First of all, notice that the ME (19) generates the exact time evolution, i.e., for any $p_y(t) = p_\beta(t) p_{y|\beta}(t) \in \mathcal{A}(t)$ we have

$$\sum_\beta V_{\alpha,\beta}(\lambda_t, t) p_\beta(t) = \sum_{x_\alpha} \sum_{\beta,y_\beta} W_{x_\alpha, y_\beta}(\lambda_t) p_{y|\beta}(t) p_\beta(t). \tag{32}$$

For the first condition, if $\pi_x(\lambda_t) \in \mathcal{A}(t) \subset \mathcal{A}_\pi(\lambda_t)$, then one immediately verifies that $V(\lambda_t, t)\boldsymbol{\pi}(\lambda_t) = 0$. But, one may have that $\mathcal{A}(t) \subset \mathcal{A}_\pi(\lambda_t)$, but $\pi_x(\lambda_t) \notin \mathcal{A}(t)$. This means that there is no admissible initial state, which gets mapped to the IFP at time t, i.e., $T_{t,0}^{-1}\boldsymbol{\pi}(\lambda_t) \notin \mathcal{A}(0)$. However, by the invertibility of the dynamics there is always a set of states $p_x^{(i)}(t) \in \mathcal{A}(t)$, which spans the entire mesostate space. Thus, we can always find a linear combination $\pi_x(\lambda_t) = \sum_i \mu_i p_x^{(i)}(t)$ with $\mu_i \in \mathbb{R}$. Then, $V(\lambda_t, t)\boldsymbol{\pi}(\lambda_t) = 0$ follows from the linearity of the dynamics by applying Eq. (32) to each term of the linear combination.

For the second condition, let us assume the opposite, i.e., $V(\lambda_t, t)\boldsymbol{\pi}(\lambda_t) \neq 0$. This implies $\sum_\beta G_{t+\delta t,t}(\alpha|\beta)\pi_\beta(\lambda_t) \neq \pi_\alpha(\lambda_t)$ for a sufficiently small δt. But, as the reduced dynamics is exact, this can only be the case if there is a state $q_y(t) = \pi_\beta(\lambda_t) q_{y|\beta}(t) \in \mathcal{A}(t)$ with $q_{y|\beta}(t) \neq \pi_{y|\beta}(\lambda_t)$. On the other hand, the theorem assumes that $\pi_x(\lambda_t) \in \mathcal{A}(t)$ too. Hence, there must be two states $q_y(t) \in \mathcal{A}(t)$ and $\pi_x(\lambda_t) \in \mathcal{A}(t)$, which give the same marginal mesostate $\pi_\alpha(\lambda_t)$. Since the ME dynamics in the full space is clearly invertible and since the initial conditional microstate is fixed, this means that there must be two different initial mesostates, which get mapped to the same mesostate at time t. Hence, $G_{t,0}$ cannot be invertible, which conflicts with our initial assumption. ■

Theorem 5 plays an important role in the limit of TSS (see Sec. III C) where the first condition is automatically fulfilled.

The second condition will be in general complicated to check if the microdynamics are complex.

It is worthwhile to ask whether milder conditions suffice to ensure that $\pi_\alpha(\lambda_t)$ is an IFP of the mesodynamics. In Appendix A we show that they can indeed be found if the dynamics fulfills the special property of weak lumpability. In general, however, we believe that it will be hard to find milder conditions: in Sec. VI A we give an example for an ergodic and undriven Markov chain, whose mesodynamics are 1-Markovian, but π_α is not an IFP unless $\mathcal{A}(0) \subset \mathcal{A}_\pi$. As any driven process takes the conditional microstates out of equilibrium, i.e., $\mathcal{A}(t) \nsubseteq \mathcal{A}_\pi(\lambda_t)$ in general, finding useful milder conditions to guarantee that $\pi_\alpha(\lambda_t)$ is an IFP seems unrealistic.

Before we proceed with the physical picture, we want to comment on a mathematical subtlety, which becomes relevant for the application considered in Sec. IV. In there, we will apply our findings from above to the case of Hamiltonian dynamics described on the continuous phase space of a collection of classical particles. This does not fit into the conventional picture of a finite and discrete state space $\mathcal{X}$ with $N < \infty$ microstates. However, under the assumption that it is possible to approximate the actual Hamiltonian dynamics by using a high-dimensional grid of very small phase space cells, we can imagine that we can approximate the true dynamics arbitrarily well with a finite, discretized phase space. Nevertheless, in order not to rely on this way of reasoning, we briefly re-derive the above theorems for the Hamiltonian setting in Appendix C.

III. COARSE-GRAINED DISSIPATIVE DYNAMICS

A. Thermodynamics at the microlevel

We now start to investigate the first application of the general framework from Sec. II. In this section, we consider the ME (14), which describes a large class of dissipative classical and quantum systems, with applications ranging from molecular motors to thermoelectric devices. In addition, we impose the condition of local detailed balance

$$\ln \frac{W_{x,y}(\lambda_t)}{W_{y,x}(\lambda_t)} = -\beta[E_x(\lambda_t) - E_y(\lambda_t)], \qquad (33)$$

where $E_x(\lambda_t)$ denotes the energy of state x and β the inverse temperature of the bath. Equation (33) ensures that the IFP at the microlevel is given by the Gibbs state $\pi_x(\lambda_t) = e^{-\beta E_x(\lambda_t)}/Z(\lambda_t)$ with $Z(\lambda_t) = \sum_x e^{-\beta E_x(\lambda_t)}$ and it allows us to link energetic changes in the system with entropic changes in the bath. A thermodynamically consistent description of the microdynamics follows from the definitions

$$U_{\text{mic}}(t) \equiv \sum_x E_x(\lambda_t) p_x(t) \quad \text{(internal energy)}, \qquad (34)$$

$$\dot{W}_{\text{mic}}(t) \equiv \sum_x [d_t E_x(\lambda_t)] p_x(t) \quad \text{(work rate)}, \qquad (35)$$

$$\dot{Q}_{\text{mic}}(t) \equiv \sum_x E_x(\lambda_t) \partial_t p_x(t) \quad \text{(heat rate)}, \qquad (36)$$

$$S_{\text{mic}}(t) \equiv -\sum_x p_x(t) \ln p_x(t) \quad \text{(Shannon entropy)}, \qquad (37)$$

$$F_{\text{mic}}(t) \equiv U_{\text{mic}}(t) - S_{\text{mic}}(t)/\beta \quad \text{(free energy)}, \qquad (38)$$

$$\dot{\Sigma}_{\text{mic}}(t) \equiv -\frac{\partial}{\partial t}\bigg|_{\lambda_t} D[p_x(t) \| \pi_x(\lambda_t)] \geqslant 0 \quad \text{(EP rate)}. \qquad (39)$$

Here, we used the subscript "mic" to emphasize that the above definitions refer to the thermodynamic description of the microdynamics, which has to be distinguished from the thermodynamic description at the mesolevel introduced below. Using the ME (14) and local detailed balance (33) together with the definitions provided above, one can verify the first and second law of thermodynamics in the conventional form $d_t U_{\text{mic}}(t) = \dot{W}_{\text{mic}}(t) + \dot{Q}_{\text{mic}}(t)$ and $\dot{\Sigma}_{\text{mic}}(t) = \beta[\dot{W}_{\text{mic}}(t) - d_t F_{\text{mic}}(t)] \geqslant 0$.

Since the IFP at the microlevel is the equilibrium Gibbs state, we can parametrize the conditional equilibrium state of the microstates belonging to a mesostate α as

$$\pi_{x|\alpha}(\lambda_t) = e^{-\beta[E_{x_\alpha}(\lambda_t) - F_\alpha(\lambda_t)]}, \qquad (40)$$

where $F_\alpha(\lambda_t) \equiv -\beta^{-1} \ln \sum_{x_\alpha} e^{-\beta E_{x_\alpha}(\lambda_t)}$ plays the role of an effective free energy. The reduced equilibrium distribution of a mesostate can then be written as

$$\pi_\alpha(\lambda_t) = \frac{e^{-\beta F_\alpha(\lambda_t)}}{Z(\lambda_t)}. \qquad (41)$$

In the following, we want to find meaningful definitions, which allow us to formulate the laws of thermodynamics at a coarse-grained level and which we can connect to the general theory of Sec. II. Since the dynamics at the mesolevel will typically be non-Markovian and not fulfill local detailed balance, finding a consistent thermodynamic framework becomes nontrivial. We will restrict our investigations here to any initial preparation class which fulfills $\mathcal{A}(0) \subset \mathcal{A}_\pi(\lambda_0)$ with $\mathcal{A}_\pi(\lambda_0)$ defined in Eq. (12). If the dynamics is driven, we will need one additional assumption [see Eq. (42)], otherwise our results are general.

B. Thermodynamics at the mesolevel

With the framework from Sec. II we are now going to study the thermodynamics at the mesolevel. This is possible in full generality if the dynamics is undriven. In case of driving, $\dot{\lambda}_t \neq 0$, we need to assume that we can split the time-dependent energy function as

$$E_{x_\alpha}(\lambda_t) = E_\alpha(\lambda_t) + \tilde{E}_{x_\alpha}. \qquad (42)$$

Thus, solely the mesostate energies are affected by the driving. This condition naturally arises if we think about the complete system as being composed of two interacting systems $\mathcal{X} = \mathcal{Y} \otimes \mathcal{Z}$, and we trace out the degrees of freedom $\mathcal{Y}$ to obtain a reduced description in $\mathcal{Z}$. In this case, we can split the energy for any value of λ_t as $E_{yz} = E_y + E_z + V_{yz}$ where V_{yz} describes an interaction energy and E_y (E_z) are the bare energies associated with the isolated system $\mathcal{Y}$ ($\mathcal{Z}$). Condition (42) is then naturally fulfilled if we identify $E_z = E_\alpha$ and only $E_z = E_z(\lambda_t)$ is time dependent (compare also with Sec. IV).

Importantly, this condition allows us to identify

$$\dot{W}_{\text{mic}}(t) = \sum_x \frac{\partial E_x(\lambda_t)}{\partial t} p_x(t)$$

$$= \sum_\alpha \frac{\partial E_\alpha(\lambda_t)}{\partial t} p_\alpha(t) \equiv \dot{W}(t). \qquad (43)$$

Therefore, the exact rate of work can be computed from the knowledge about the mesostate alone. Furthermore, Eq. (42) implies that the conditional equilibrium state of the bath (40) does not depend on λ_t and, hence, we can write $\mathcal{A}_\pi(\lambda_0) = \mathcal{A}_\pi$.

The thermodynamic analysis starts from our central definition (24)

$$\dot{\Sigma}(t) = -\left.\frac{\partial}{\partial t}\right|_{\lambda_t} D[p_\alpha(t)\|\pi_\alpha(\lambda_t)] \qquad (44)$$

with $\pi_\alpha(\lambda_t)$ given in Eq. (41). Using Eq. (43) and noting that $d_t F_\alpha(\lambda_t) = d_t E_\alpha(\lambda_t)$, it is not hard to confirm that

$$\dot{\Sigma}(t) = \beta \dot{W}(t) - \beta \frac{d}{dt} \sum_\alpha p_\alpha(t) \left[F_\alpha(\lambda_t) + \frac{1}{\beta} \ln p_\alpha(t) \right]. \qquad (45)$$

This motivates the definition of the nonequilibrium free energy

$$F(t) \equiv \sum_\alpha p_\alpha(t) \left[F_\alpha(\lambda_t) + \frac{1}{\beta} \ln p_\alpha(t) \right], \qquad (46)$$

such that the EP rate is given by the familiar form of phenomenological nonequilibrium thermodynamics: $\dot{\Sigma}(t) = \beta[\dot{W}(t) - d_t F(t)]$. The EP over a finite time interval becomes

$$\Sigma(t) = \beta[W(t) - \Delta F(t)] \qquad (47)$$

and for a proper second law it remains to show that this quantity is positive. This follows from the next theorem:

Theorem 6. For any $p_x(0) \in \mathcal{A}_\pi$ and any driving protocol we have

$$\Sigma(t) \geqslant \Sigma_{\text{mic}}(t) \geqslant 0. \qquad (48)$$

Proof. The proof was already given in Ref. [32]. In short, one rewrites

$$\Sigma(t) - \Sigma_{\text{mic}}(t) = \beta[\Delta F_{\text{mic}}(t) - \Delta F(t)] \qquad (49)$$

and shows that for $p_x(0) \in \mathcal{A}_\pi$ it follows that

$$\beta[\Delta F_{\text{mic}}(t) - \Delta F(t)]$$
$$= D[p_x(t)\|\pi_x(\lambda_t)] - D[p_\alpha(t)\|\pi_\alpha(\lambda_t)]$$
$$= \sum_\alpha p_\alpha(t) D[p_{x|\alpha}(t)\|\pi_{x|\alpha}] \geqslant 0. \qquad (50)$$

Since $\Sigma_{\text{mic}}(t) \geqslant 0$, this implies $\Sigma(t) \geqslant 0$. ∎

Using the theorems of Sec. II B, we can now connect the appearance of negative EP rates to the following properties of the underlying dynamics:

Theorem 7. Let $p_x(0) \in \mathcal{A}_\pi$ and let I denote the time interval in which the mesodynamics are 1-Markovian and the dynamics is

(1) undriven, or
(2) driven and lumpable, or

(3) driven and such that $\mathcal{A}(t) \subset \mathcal{A}_\pi$ or $\pi_x(\lambda_t) \in \mathcal{A}(t)$.

Then, $\dot{\Sigma}(t) \geqslant 0$ for all $t \in I$ and all admissible initial states.

Hence, as a corollary, if we observe $\dot{\Sigma}(t) < 0$ for the undriven case, we know that the dynamics is non-Markovian [or that the initial state $p_x(0) \notin \mathcal{A}_\pi$]. For driven dynamics, noticing a negative EP rate is not sufficient to conclude that the dynamics is non-Markovian, but it is clearly not lumpable. In the next section, we will show that $\dot{\Sigma}(t) < 0$ also suffices to conclude that TSS does not apply.

Furthermore, while the above procedure provides a unique way to define a nonequilibrium free energy at the mesolevel, it does not fix the definition of the internal energy and entropy at the mesolevel because the prescription $F = U - S/\beta$ entails a certain level of arbitrariness. Via the first law $\Delta U = Q + W$ this would also imply a certain arbitrariness for the definition of heat [44]. However, a reasonable definition of U, S, and Q can be fixed by demanding that they should coincide with U_{mic}, S_{mic}, and Q_{mic} in the limit where the microstates are conditionally equilibrated, which is fulfilled in the limit of TSS considered in Sec. III C. Then, one is naturally led to the definitions

$$U(t) \equiv \sum_\alpha \mathcal{U}_\alpha(\lambda_t) p_\alpha(t), \quad \mathcal{U}_\alpha \equiv \sum_{x_\alpha} E_{x_\alpha}(\lambda_t) \pi_{x|\alpha}, \qquad (51)$$

$$S(t) \equiv \sum_\alpha \{\beta[\mathcal{U}_\alpha(\lambda_t) - F_\alpha(\lambda_t)] - \ln p_\alpha(t)\} p_\alpha(t). \qquad (52)$$

Heat is then defined as $\dot{Q}(t) = d_t U(t) - \dot{W}(t)$ and the EP rate can be equivalently expressed as $\dot{\Sigma}(t) = d_t S(t) - \beta \dot{Q}(t)$.

We remark that it is not obvious how to relax condition (42) because the work (43) can then not be computed from knowledge of the mesostate alone, which was an essential ingredient in our derivation.

C. Timescale separation and Markovian limits

Although open systems behave non-Markovian in general, it is important to know in which limits the Markovian *approximation* is justified. One such limit is TSS, which is an essential assumption in many branches of statistical mechanics in order to ensure that the dynamics at the level of the "relevant" degrees of freedom is Markovian and, hence, easily tractable. It is also essential in order to ensure that we can infer from the coarse-grained dynamics the *exact* thermodynamics of the underlying microstate dynamics (under reasonable mild conditions) (see Refs. [22,32,45–48] for research on this topic). Here, we restrict ourselves to highlight the role of TSS within our mathematical framework of Sec. II. Furthermore, at the end of this section we discuss another class of systems whose dynamics is Markovian, albeit TSS does not apply.

To study TSS, let us decompose the rate matrix as follows:

$$W_{x_\alpha, y_\beta}(\lambda_t) = \delta_{\alpha\beta} R_{x_\alpha, y_\alpha}(\lambda_t) + (1 - \delta_{\alpha\beta}) r_{x_\alpha, y_\beta}(\lambda_t). \qquad (53)$$

Next, we assume that $R_{x_\alpha, y_\alpha}(\lambda_t) \gg r_{x_\alpha, y_\beta}(\lambda_t)$, i.e., there is a strong separation of timescales between the mesodynamics and the microdynamics belonging to a certain mesostate. As a consequence, the microstates rapidly equilibrate to the conditional steady state $\pi_{x|\alpha}(\lambda_t)$ for any mesostate α provided that the microstates in each mesostate are fully connected

(tacitly assumed in the following). This means that condition 1 of Theorem 5 is always fulfilled. By replacing $p_{y|\beta}(t)$ by $\pi_{y|\beta}(\lambda_t)$ in Eq. (17), it is easy to see that the effective rate matrix is independent of the initial state and describes a proper Markov process $R[\lambda_t, p_\alpha(0)] = R(\lambda_t)$. Another consequence of TSS is that the thermodynamics associated with the mesodynamics are identical to the thermodynamics of the microdynamics.

Strictly speaking, the limit of TSS requires $R_{x_\alpha, y_\alpha}(\lambda_t)/r_{x_\alpha, y_\beta}(\lambda_t) \to \infty$. In practice, however, there will be always a finite time δt associated with the relaxation of the microstates and TSS means that we assume

$$\frac{1}{r_{x_\alpha, y_\beta}(\lambda_t)} \gg \delta t \gg \frac{1}{R_{x_\alpha, y_\alpha}(\lambda_t)}. \quad (54)$$

Then, within a time step δt the conditional microstates are almost equilibrated while terms of the order $O(\delta t^2 r_{x_\alpha, y_\beta})$ are still negligible. The TM in this situation becomes

$$T_{t+\delta t, t}(x_\alpha | y_\beta) \approx \delta_{\alpha\beta} \pi_{x|\alpha}(\lambda_t) \left(1 - \delta t \sum_{\gamma \neq \alpha} \sum_{z_\gamma} r_{z_\gamma, x_\alpha}(\lambda_t) \right)$$
$$+ \delta t (1 - \delta_{\alpha\beta}) \sum_{z_\beta} \pi_{z|\beta}(\lambda_t) r_{x_\alpha, z_\beta}(\lambda_t). \quad (55)$$

The first term describes the probability for a transition within two microstates of the same mesostate: to lowest order, this is simply given by the conditional steady state minus a small correction term of $O(\delta t)$, which takes into account the possibility that one leaves the given mesostate to another mesostate. The second term gives the probability to reach a microstate lying in a different mesostate, which is given by the sum of all possible rates which connect to this microstate from the given mesostate multiplied by the respective conditional steady state probability. One immediately checks normalization of $T_{t+\delta t, t}(x_\alpha | y_\beta)$ and positivity follows by assuming that $r_{z_\gamma, x_\alpha}(\lambda_t) \delta t \ll 1$. Furthermore, also the condition (21) of lumpability is fulfilled. Indeed, we can even confirm the stronger property

$$T_{t+\delta t, t}(x_\alpha | y_\beta) = T_{t+\delta t, t}(x_\alpha | y'_\beta) \quad (56)$$

for all $y'_\beta \neq y_\beta$. Hence, in the idealized limit yielding to an instantaneous equilibration of the conditional microstates, the TMs do not even depend on the particular microstate anymore. We conclude as follows:

Theorem 8. If TSS applies, then the process is lumpable and $p_x(t) \in \mathcal{A}_\pi$ for all t. Conversely, if $\dot{\Sigma}(t) < 0$, then TSS does not apply.

It was shown in Ref. [22] that $\dot{\Sigma}(t) = \dot{\Sigma}_{\text{mic}}(t)$ in the limit of TSS. If only the slightly weaker condition of lumpability is fulfilled, then it is not known whether $\dot{\Sigma}(t) = \dot{\Sigma}_{\text{mic}}(t)$ still holds.

While TSS is an important limit, the mesodynamics can be also Markovian without the assumption of TSS. The following theorem demonstrates this explicitly:

Theorem 9. If there is a partition χ such that the rate matrix can be written as

$$W_{x_\alpha, y_\beta}(\lambda_t) = \delta_{\alpha\beta} R_{x_\alpha, y_\alpha}(\lambda_t) + (1 - \delta_{\alpha\beta}) V_{\alpha, \beta}(\lambda_t), \quad (57)$$

then the process is lumpable independent of any TSS argument. Moreover, the IFP of the lumped process is $\pi_\alpha(\lambda_t) = \sum_{x_\alpha} \pi_{x_\alpha}(\lambda_t)$ and, hence, $\dot{\Sigma}(t) \geqslant 0$ always.

Proof. We first of all observe that from

$$0 = \sum_{\alpha, x_\alpha} W_{x_\alpha, y_\beta}(\lambda_t)$$
$$= \sum_{x_\beta} R_{x_\beta, y_\beta}(\lambda_t) + \sum_{\alpha \neq \beta} \sum_{x_\alpha} V_{\alpha, \beta}(\lambda_t), \quad (58)$$

it follows that $\sum_{x_\alpha} R_{x_\alpha, y_\alpha}(\lambda_t) = -\sum_{\beta \neq \alpha} \#\chi_\alpha V_{\beta, \alpha}(\lambda_t)$ for any α (where $\#\chi_\alpha$ denotes the cardinality of the set of microstates belonging to mesostate α). By using this property, it becomes straightforward to check that Eq. (22) is fulfilled and, hence, the coarse-grained process is Markovian. Due to Theorem 3 we can also confirm that $\pi_\alpha(\lambda_t)$ is the IFP and from Theorem 2 it follows that $\dot{\Sigma}(t) \geqslant 0$. ∎

Compared to the decomposition (53) we here did not need to assume any particular scaling of the rates, but it was important that the transitions between different mesostates are independent of the microstate. In fact, for many mesoscopic systems, the details of the microstates might not matter, for instance, the Brownian motion of a suspended particle is quite independent from the spin degrees of freedom of its electrons unless strong magnetic interactions are present. Notice that the ME at the mesolevel resulting from Eq. (57) reads as

$$\frac{\partial}{\partial t} p_\alpha(t) = \sum_{\beta \neq \alpha} [\#\chi_\alpha V_{\alpha, \beta}(\lambda_t) p_\beta(t) - \#\chi_\beta V_{\beta, \alpha}(\lambda_t) p_\alpha(t)].$$
$$(59)$$

It shows that the local detailed balance ratio (33) of the effective rates at the mesolevel is shifted by an entropic contribution due to the degeneracy factor $\#\chi_\alpha$; see Sec. VI B or Ref. [49] for explicit examples.

IV. CLASSICAL SYSTEM-BATH THEORY

In this section we consider the standard paradigm of classical open system theory: a system in contact with a bath described by Hamiltonian dynamics as opposed to the rate ME dynamics from Sec. III. The microstates (system and bath) therefore describe an isolated system and the goal is to find a consistent thermodynamic framework for the mesostate (the system only). The global Hamiltonian reads as

$$H_{\text{tot}}(\lambda_t) = H(\lambda_t) + V + H_B, \quad (60)$$

where the system, bath, and interaction Hamiltonian $H(\lambda_t)$, H_B, and V are arbitrary. We denote a phase space point of the system by x_S and of the bath by x_B. Thus, to be very precise, we should write $H(x_S; \lambda_t)$, $H_B(x_B)$, and $V(x_S, x_B)$, but we will drop the dependency on x_S and x_B for notational simplicity. Deriving the laws of thermodynamics for an arbitrary Hamiltonian (60) has attracted much interest recently [29–32,44,50–52] (note that many investigations in the quantum domain also have a direct analog in the classical regime [53–59]). It will turn out that our basic definitions are identical to the ones suggested by Seifert [29]. We here rederive them in a different way and, in addition, we focus on the EP *rate* and its relation to non-Markovian dynamics.

In order to be able to define the EP rate (24), we first of all need to know the exact equilibrium state of the system, which is obtained from coarse graining the global equilibrium state $\pi_{\text{tot}}(\lambda_t) = e^{-\beta H_{\text{tot}}(\lambda_t)}/\mathcal{Z}_{\text{tot}}(\lambda_t)$ with $\mathcal{Z}_{\text{tot}}(\lambda_t) = \int dx_{SB} e^{-\beta H_{\text{tot}}(\lambda_t)}$. For this purpose we introduce the Hamiltonian of mean force $H^*(\lambda_t)$ [60]. It is defined through the two relations

$$\pi_S(\lambda_t) \equiv \frac{e^{-\beta H^*(\lambda_t)}}{\mathcal{Z}^*(\lambda_t)} = \int dx_B \frac{e^{-\beta H_{\text{tot}}(\lambda_t)}}{\mathcal{Z}_{\text{tot}}(\lambda_t)},$$

$$\mathcal{Z}^*(\lambda_t) \equiv \frac{\mathcal{Z}_{\text{tot}}(\lambda_t)}{\mathcal{Z}_B}, \quad (61)$$

where $\mathcal{Z}_B = \int dx_B e^{-\beta H_B}$ is the equilibrium partition function of the unperturbed bath. We emphasize that the equilibrium state of the system is not a Gibbs state with respect to $H(\lambda_t)$ due to the strong coupling. More explicitly, the Hamiltonian of mean force reads as

$$H^*(\lambda_t) = H(\lambda_t) - \frac{1}{\beta} \ln \langle e^{-\beta V} \rangle_B^{\text{eq}}, \quad (62)$$

where $\langle \ldots \rangle_B^{\text{eq}}$ denotes an average with respect to the unperturbed equilibrium state of the bath $e^{-\beta H_B}/\mathcal{Z}_B$. Note that $H^*(\lambda_t)$ also depends on the inverse temperature β of the bath.

We can now use Eq. (24) to define the EP rate, which reads as in the notation of this section

$$\dot{\Sigma}(t) = -\frac{\partial}{\partial t}\bigg|_{\lambda_t} D[\rho_S(t) \| \pi_S(\lambda_t)], \quad (63)$$

where $\rho_S(t) = \rho_S(x_S; t)$ denotes the state of the system at time t, which can be arbitrarily far from equilibrium. Note that we now use the differential relative entropy $D[\rho_S(t) \| \pi_S(\lambda_t)] = \int dx_S \rho_S(x_S; t) \ln \frac{\rho_S(x_S;t)}{\pi_S(x_S;\lambda_t)}$. Using Eq. (61), we can rewrite Eq. (63) as

$$\dot{\Sigma}(t) = \frac{d}{dt} S[\rho_S(t)] - \beta \int dx_S H^*(\lambda_t) \frac{d}{dt} \rho_S(t) \quad (64)$$

with $S[\rho_S(t)] \equiv -\int dx_S \rho(x_S; t) \ln \rho(x_S; t)$. The second term can be cast into the form

$$\int dx_S H^*(\lambda_t) \frac{d}{dt} \rho_S(t) = \frac{d}{dt} \langle H^*(\lambda_t) \rangle - \left\langle \frac{dH^*(\lambda_t)}{dt} \right\rangle, \quad (65)$$

where $\langle \ldots \rangle$ denotes a phase space average with respect to $\rho_S(t)$. After realizing that $d_t H^*(\lambda_t) = d_t H(\lambda_t)$, we see that the last term coincides with the rate of work done on the system

$$\dot{W}(t) = \int dx_S \frac{dH(\lambda_t)}{dt} \rho_S(t). \quad (66)$$

Using

$$\int dx_S \frac{dH(\lambda_t)}{dt} \rho_S(t) = \int dx_{SB} \frac{dH_{\text{tot}}(\lambda_t)}{dt} \rho_{\text{tot}}(t)$$

$$= \int dx_{SB} \frac{d}{dt} [H_{\text{tot}}(\lambda_t) \rho_{\text{tot}}(t)], \quad (67)$$

this can be integrated to

$$W(t) = \int_0^t ds \left\langle \frac{dH(\lambda_s)}{ds} \right\rangle$$

$$= \int dx_{SB} [H_{\text{tot}}(\lambda_t) \rho_{\text{tot}}(t) - H_{\text{tot}}(\lambda_0) \rho_{\text{tot}}(0)], \quad (68)$$

showing that the work done on the system is given by the total energetic change of the composite system and environment. The EP rate can then be expressed as

$$\dot{\Sigma}(t) = \beta \left[\dot{W}(t) - \frac{d}{dt} \left\langle H^*(\lambda_t) + \frac{1}{\beta} \ln \rho_S(t) \right\rangle \right]. \quad (69)$$

This motivates again the following definition of the nonequilibrium free energy [cf. Eq. (46)]

$$F(t) \equiv \left\langle H^*(\lambda_t) + \frac{1}{\beta} \ln \rho_S(t) \right\rangle \quad (70)$$

such that $\dot{\Sigma}(t) = \beta[\dot{W}(t) - d_t F(t)]$.

For a useful thermodynamic framework, it now remains to show that the second law as known from phenomenological nonequilibrium thermodynamics holds:

$$\Sigma(t) \equiv \beta[W(t) - \Delta F(t)] \geqslant 0. \quad (71)$$

For this purpose, we assume as in the previous section that the initial state $\rho_S(0)$ belongs to the set $\mathcal{A}_\pi$ [see Eq. (12)]. The conditional equilibrium state of the bath is given by

$$\pi_{B|S} \equiv \frac{e^{-\beta(V + H_B)}}{\int dx_B e^{-\beta(V + H_B)}} = \frac{e^{-\beta[H_{\text{tot}}(\lambda_0) - H(\lambda_0)]}}{\mathcal{Z}_B}. \quad (72)$$

To prove the positivity of the EP, we refer to Ref. [29], where it was deduced from an integral fluctuation theorem or, alternatively, the positivity becomes evident by noting the relation $\Sigma(t) = D[\rho_{SB}(t) \| \rho_S(t) \pi_{B|S}]$ and by recalling that the relative entropy is always positive [31,32]. It is important to realize, however, that $\Sigma(t) \geqslant 0$ relies crucially on the choice of initial state. If $\rho_{\text{tot}}(0) \notin \mathcal{A}_\pi$, we have

$$\beta[W(t) - \Delta F(t)] = D[\rho_{SB}(t) \| \rho_S(t) \pi_{B|S}]$$

$$- D[\rho_{SB}(0) \| \rho_S(0) \pi_{B|S}], \quad (73)$$

which can be negative.

After we have established that $\Sigma(t) = \int_0^t ds \dot{\Sigma}(s) \geqslant 0$ with the EP rate $\dot{\Sigma}(t)$ from Eq. (24), we can use the insights from Sec. II and Appendix C. Then, we can immediately confirm the validity of the following theorem:

Theorem 10. Let $\rho_{\text{tot}}(0) \in \mathcal{A}_\pi$ and let I denote the time interval in which the system dynamics is 1-Markovian and the process is

   (1) undriven, or

   (2) driven and lumpable, or

   (3) driven and $\mathcal{A}(t) \subset \mathcal{A}_\pi$ or $\pi_{\text{tot}}(\lambda_t) \in \mathcal{A}(t)$.

Then, $\dot{\Sigma}(t) \geqslant 0$ for all $t \in I$ and all admissible initial states.

We can therefore conclude for this setup that $\dot{\Sigma}(t) < 0$ directly implies non-Markovian dynamics for *undriven* systems. For *driven* systems this relation ceases to exist, but similar to Theorem 8, $\dot{\Sigma}(t) < 0$ implies that the two assumptions of 1-Markovian dynamics and a bath in a conditional equilibrium

state cannot be simultaneously fulfilled. Two further remarks are in order:

First, although it is possible to extend the framework of Ref. [29] to the situation of a time-dependent coupling Hamiltonian $V(\lambda_t)$ (see Ref. [32]), Theorem 10 then ceases to hold because the work (68) cannot anymore be computed from knowledge of the system state alone [also compare with Eq. (43)].

Second, we remark that Theorem 10 is structurally identical to Theorem 7. This shows the internal consistency of our approach: since it is in principle possible to derive a ME from underlying Hamiltonian dynamics, we should find parallel results at each level of the description. This structural similarity was also found in Ref. [32].

Also in parallel to Sec. III, we remark that the splitting of the free energy $F = U - S/\beta$ does not allow to unambiguously define an internal energy and entropy. Hence, also the definition of heat via the first law $\Delta U = Q + W$ becomes ambiguous [44]. However, the following definitions are appealing:

$$U(t) \equiv \int dx_S \rho_S(t)[H^*(\lambda_t) + \beta\partial_\beta H^*(\lambda_t)], \quad (74)$$

$$S(t) \equiv \int dx_S \rho_S(t)[-\ln \rho_S(t) + \beta^2\partial_\beta H^*(\lambda_t)], \quad (75)$$

which can be shown to coincide (apart from a time-independent additive constant) with the global energy and entropy in equilibrium [29]. Further support for these definitions was given in Ref. [32]; see also the discussion in Ref. [30].

Finally, to gain further insights into our approach, it is useful to reformulate it in terms of expressions which were previously derived for classical Hamiltonian dynamics [61–65]. It follows from straightforward algebra that

$$D[\rho_{\text{tot}}(t)\|\pi_{\text{tot}}(\lambda_t)] = \beta[F_{\text{tot}}(t) - \mathcal{F}_{\text{tot}}(\lambda_t)], \quad (76)$$

where $F_{\text{tot}}(t) = \langle H_{\text{tot}}(\lambda_t)\rangle + \langle\ln \rho_{\text{tot}}(t)\rangle/\beta$ is the nonequilibrium free energy associated to the global state $\rho_{\text{tot}}(t)$ and $\mathcal{F}_{\text{tot}}(\lambda_t)$ is the equilibrium free energy associated to the thermal state $\pi_{\text{tot}}(\lambda_t)$. Due to Eq. (76) we can write the global EP rate as

$$\dot{\Sigma}_{\text{tot}}(t) = -\frac{\partial}{\partial t}\bigg|_{\lambda_t} D[p_x(t)\|\pi_x(\lambda_t)]$$

$$= \beta\left[\dot{W}_{\text{irr}}(t) - \frac{d}{dt}D[p_x(t)\|\pi_x(\lambda_t)]\right] = 0, \quad (77)$$

which is zero for Hamiltonian dynamics. Here, $\dot{W}_{\text{irr}}(t) \equiv \dot{W} - d_t\mathcal{F}_{\text{tot}}(\lambda_t)$ is the irreversible work and, thus, Eq. (77) recovers (parts of) the earlier results from Refs. [61–65]. Especially for an initially equilibrated microstate we immediately get the well-known dissipation inequality $W_{\text{irr}}(t) = D[p_x(t)\|\pi_x(\lambda_0)] \geqslant 0$. Now, from our findings above we see that we obtain an identical structure at the coarse-grained level: by using the identity (76) for the system $D[\rho_S(t)\|\pi_S(\lambda_t)] = \beta[F(t) - \mathcal{F}(\lambda_t)]$, we obtain

$$\dot{\Sigma}(t) = \beta\left[\dot{W}_{\text{irr}}(t) - \frac{d}{dt}D[p_\alpha(t)\|\pi_\alpha(\lambda_t)]\right]. \quad (78)$$

This expression can in general be negative and the conditions which ensure non-negativity are stated in Theorem 10.

V. STRONG COUPLING THERMODYNAMICS OF QUANTUM SYSTEMS

So far, we have only treated classical systems, but the question of how to obtain a meaningful thermodynamic description for quantum systems beyond the weak coupling and Markovian approximation is of equal importance. Whereas in Sec. IV we could resort to an already well-developed framework, no general finite-time thermodynamic description for a driven quantum system immersed in an arbitrary single heat bath has been presented yet. Based on results obtained at equilibrium [51,66], we first of all develop in Sec. V A the quantum extension of the framework introduced in Ref. [29]. Afterwards, in Sec. V B we prove that the relation worked out between non-Markovianity and a negative EP rate for classical systems cannot be established for quantum systems. The latter point is further studied in Sec. VI D for the commonly used assumption that the system and bath are initially decorrelated, an assumption which is not true for the class of initial states considered in this section.

A. Integrated description

As in Sec. IV our starting point is a time-dependent system-bath Hamiltonian of the form $\hat{H}_{\text{tot}}(\lambda_t) = \hat{H}(\lambda_t) + \hat{V} + \hat{H}_B$, where we used a hat to explicitly denote operators. The Hamiltonian of mean force in the quantum case is formally given by

$$\hat{H}^*(\lambda_t) = -\frac{1}{\beta}\ln\frac{\text{tr}_B\{e^{-\beta[\hat{H}(\lambda_t)+\hat{V}+\hat{H}_B]}\}}{Z_B} \quad (79)$$

and it shares the same meaning as in the classical case [cf. Eq. (61)]: it describes the exact reduced state of the system if the system-bath composite is in a global equilibrium state. Motivated by equilibrium considerations and by Sec. IV, we define the three key thermodynamic quantities, internal energy, system entropy, and free energy, for an arbitrary system state $\hat{\rho}_S(t)$ as follows:

$$U(t) \equiv \text{tr}_S\{\hat{\rho}_S(t)[\hat{H}^*(\lambda_t) + \beta\partial_\beta\hat{H}^*(\lambda_t)]\}, \quad (80)$$

$$S(t) \equiv \text{tr}_S\{\hat{\rho}_S(t)[-\ln\hat{\rho}_S(t) + \beta^2\partial_\beta\hat{H}^*(\lambda_t)]\}, \quad (81)$$

$$F(t) \equiv \text{tr}_S\left\{\hat{\rho}_S(t)\left[\hat{H}^*(\lambda_t) + \frac{1}{\beta}\ln\hat{\rho}_S(t)\right]\right\}. \quad (82)$$

Note that all quantities are state functions. Also, the definition of work is formally identical to Sec. IV, Eq. (68):

$$W(t) = \int_0^t ds\,\text{tr}_S\left\{\frac{d\hat{H}(\lambda_s)}{ds}\hat{\rho}_S(s)\right\}$$

$$= \text{tr}_{SB}\{\hat{\rho}_{\text{tot}}(t)\hat{H}_{\text{tot}}(\lambda_t)\} - \text{tr}_{SB}\{\hat{\rho}_{\text{tot}}(0)\hat{H}_{\text{tot}}(\lambda_0)\}, \quad (83)$$

and the heat flux is again fixed by the first law $Q(t) = \Delta U(t) - W(t)$.

Equipped with these definitions, we define the EP

$$\Sigma(t) \equiv \beta[W(t) - \Delta F(t)] \quad (84)$$

as usual and ask when can we ensure its positivity? Again, in complete analogy to Eq. (73) one can show that

$$
\begin{aligned}
\beta[W(t) &- \Delta F(t)] \\
&= D[\hat{\rho}_{\text{tot}}(t)\|\hat{\pi}_{\text{tot}}(\lambda_t)] - D[\hat{\rho}_S(t)\|\hat{\pi}_S(\lambda_t)] \\
&\quad - D[\hat{\rho}_{\text{tot}}(0)\|\hat{\pi}_{\text{tot}}(\lambda_0)] + D[\hat{\rho}_S(0)\|\hat{\pi}_S(\lambda_0)],
\end{aligned}
\tag{85}
$$

where $D[\hat{\rho}\|\hat{\sigma}] \equiv \text{tr}\{\hat{\rho}(\ln\hat{\rho} - \ln\hat{\sigma})\}$ is the quantum relative entropy and $\hat{\pi}_{\text{tot}}(\lambda_t)$ the global Gibbs state and $\hat{\pi}_S(\lambda_t) = \text{tr}_B\{\hat{\pi}_{\text{tot}}(\lambda_t)\}$. Equation (85) can be derived by using that the von Neumann entropy of the global state $S[\hat{\rho}_{\text{tot}}(t)] = -\text{tr}_{SB}\{\hat{\rho}_{\text{tot}}(t)\ln\hat{\rho}_{\text{tot}}(t)\}$ is conserved and by using the relation $\ln[\frac{\mathcal{Z}^*(\lambda_t)}{\mathcal{Z}_{\text{tot}}(\lambda_t)}\frac{\mathcal{Z}_{\text{tot}}(\lambda_0)}{\mathcal{Z}^*(\lambda_0)}] = \ln\frac{\mathcal{Z}_B}{\mathcal{Z}_B} = 0$, where the partition functions are defined analogously to Eq. (61). Notice that this identity requires the bath Hamiltonian to be undriven.

We now note that due to the monotonicity of relative entropy [67,68], the first line in Eq. (85) is never negative, while the second line is never positive. Hence, positivity of the EP (84) is ensured if

$$
D[\hat{\rho}_{\text{tot}}(0)\|\hat{\pi}_{\text{tot}}(\lambda_0)] - D[\hat{\rho}_S(0)\|\hat{\pi}_S(\lambda_0)] = 0.
\tag{86}
$$

Two important classes of initial states for which this is the case are as follows:

Class 1 (global Gibbs state). If the initial composite system-bath state is a Gibbs state $\hat{\pi}_{\text{tot}}(\lambda_0)$, we immediately see that Eq. (86) is fulfilled and $\beta[W(t) - \Delta F(t)] \geqslant 0$ holds true. For a cyclic process, in which the system Hamiltonian is the same at the initial and final times, positivity of Eq. (84) follows alternatively from the approach in Ref. [69].

Class 2 (commuting initial state). We consider initial states of the form

$$
\hat{\rho}_{\text{tot}}(0) = \sum_k p_k(0)\hat{\Pi}_k\hat{\rho}_{B|k}(\lambda_0),
\tag{87}
$$

where the $\hat{\Pi}_k = |k\rangle\langle k|$ are orthogonal rank-1 projectors in the system space fulfilling the commutation relations

$$
[\hat{\Pi}_k, \hat{H}^*(\lambda_0)] = [\hat{\Pi}_k, \hat{H}_{\text{tot}}(\lambda_0)] = 0 \ \forall \ k.
\tag{88}
$$

This is ensured when $[\hat{H}(\lambda_0), \hat{V}] = 0$. The state of the bath conditioned on the system state $\hat{\Pi}_k$ reads as

$$
\hat{\rho}_{B|k}(\lambda_0) = \frac{\text{tr}_S\{\hat{\Pi}_k\hat{\pi}_{\text{tot}}(\lambda_0)\}}{\text{tr}_{SB}\{\hat{\Pi}_k\hat{\pi}_{\text{tot}}(\lambda_0)\}} = \frac{\langle k|\hat{\pi}_{\text{tot}}(\lambda_0)|k\rangle}{\langle k|\hat{\pi}_S(\lambda_0)|k\rangle}.
\tag{89}
$$

Since the $p_k(0)$ are allowed to be arbitrary probabilities, Eq. (87) is the direct quantum analog of the initial states considered in the classical setting in Sec. IV. Using condition (88), it becomes a task of straightforward algebra to show that Eq. (86) holds.

We remark that all considerations above can be also extended to a time-dependent coupling Hamiltonian, i.e., by allowing $\hat{V} = \hat{V}(\lambda_t)$ to depend on time. Again, the problem is then that the work (83) cannot be computed based on the knowledge of the system state $\hat{\rho}_S(t)$ alone. Furthermore, it is worth to point out that positivity of the second law (84) with the *nonequilibrium* free energy represents a stronger inequality than the bound for the dissipated work derived in Ref. [70] from a fluctuation theorem using the equilibrium free energy.

B. Breakdown of the results from Sec. IV

The positivity of $\Sigma(t)$ could be established for initial global Gibbs states or for commuting initial states. Without any driving ($\dot{\lambda}_t = 0$), these states are not very interesting as they remain invariant in time. Hence, we only consider the driven situation. Clearly, the analog of Eq. (84) at the rate level is $\beta[\dot{W}(t) - d_t F(t)]$. Unfortunately, this does not coincide with the quantum counterpart of Eq. (24). To see this, suppose that

$$
\dot{\Sigma}(t) = -\left.\frac{\partial}{\partial t}\right|_{\lambda_t} D[\hat{\rho}_S(t)\|\hat{\pi}_S(\lambda_t)].
\tag{90}
$$

This can be rewritten as

$$
\dot{\Sigma}(t) = \frac{d}{dt}\{S[\hat{\rho}_S(t)] - \beta\langle\hat{H}^*(\lambda_t)\rangle\} + \beta\text{tr}\left\{\hat{\rho}_S(t)\frac{d\hat{H}^*(\lambda_t)}{dt}\right\}.
\tag{91}
$$

Unfortunately, the analogy with Sec. IV stops here because the last term cannot be identified with the work done on the quantum system and, hence, $\int_0^t ds\,\dot{\Sigma}(s) \neq \Sigma(t)$. In fact,

$$
\frac{\partial\hat{H}^*(\lambda_t)}{\partial t} \neq \frac{\partial\hat{H}(\lambda_t)}{\partial t}
\tag{92}
$$

unless in the "classical" (and for us uninteresting) limit $[H(\lambda_t), V] = 0$.

To conclude, for quantum systems the EP rate cannot be expressed in terms of a relative entropy describing the irreversible relaxation to the equilibrium state, which would be desirable because an analog of Lemma 1 holds also in the quantum case [18]. Thus, the very existence of a general relation between EP and non-Markovianity as established for previous setups seems questionable at the moment. This conclusion can be drawn without touching upon the difficult question of how to extend many of the mathematical results of Sec. II to the quantum case.

VI. APPLICATIONS

After having established the general theory in the last four sections, we now consider various examples and applications. However, it is not our intention here to cover every aspect of our theory. We rather prefer to focus on simple models, whose essence is easy to grasp and which illuminate certain key aspects of our framework, thereby also shedding light on some misleading statements made in the literature.

A. Time-dependent instantaneous fixed points for an undriven ergodic Markov chain

For the formal development of our theory, it was of crucial importance to know under which conditions we could ensure that there is a well-defined IFP $\pi_\alpha(\lambda_t)$ for the coarse-grained dynamics, which follows from an underlying steady state of the microdynamics. Especially for driven systems this was hard to establish because even when we start with the initial steady state $\pi_x(\lambda_0)$, the driving will take it out of that state such that $p_x(t) \neq \pi_x(\lambda_t)$ in general. One might wonder whether additional conditions, such as 1-Markovianity or ergodicity, help to ensure that $\pi_\alpha(\lambda_t)$ is an IFP of the mesodynamics, but we will here show that this is not the case.
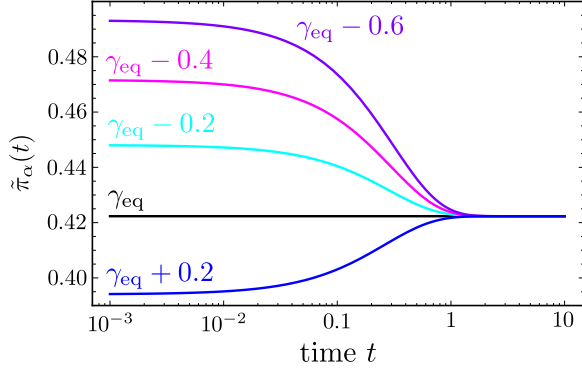
FIG. 4. Plot of the changing IFP, denoted here by $\tilde{\pi}_\alpha(t)$, over time t in logarithmic scale (for the plot we set the initial time $t^* = 0$). The figure shows that $\tilde{\pi}_\alpha(t) \neq \pi_\alpha(\lambda_{t^*})$ unless we choose $\gamma = \gamma_{\text{eq}}$. In the long-time limit, the IFP coincides with the equilibrium distribution (94). We set $\epsilon = 1$ which implies $\gamma_{\text{eq}} \approx 0.73$.

As a counterexample, we consider a simple three-state system described by a 3×3 rate matrix $W(\lambda_t)$. Imagine that the system started in $\mathcal{A}(0) \subset \mathcal{A}_\pi(\lambda_0)$, i.e., the initial microstates were conditionally equilibrated. The system is then subjected to an arbitrary driving protocol λ_t up to some time t^*. Afterwards, we keep the protocol fixed, i.e., $\lambda_t = \lambda_{t^*}$ for all $t \geq t^*$. Clearly, at time t^* the microstates will in general not be conditionally equilibrated, i.e., $\mathcal{A}(t) \nsubseteq \mathcal{A}_\pi(\lambda_{t^*})$.

Now, for definiteness we choose the full rate matrix describing the evolution of the probability vector $\mathbf{p}(t) = [p_1(t), p_2(t), p_3(t)]$ for $t \geq t^*$ to be

$$W(\lambda_{t^*}) = \begin{pmatrix} -1 - e^{-\epsilon/2} & 1 & e^{\epsilon/2} \\ 1 & -1 - e^{-\epsilon/2} & e^{\epsilon/2} \\ e^{-\epsilon/2} & e^{-\epsilon/2} & -2e^{\epsilon/2} \end{pmatrix}. \quad (93)$$

It obeys local detailed balance (33) if we parametrize the inverse temperature and energies as $\beta E_1 = \beta E_2 = 0$ and $\beta E_3 = \epsilon$ and, furthermore, we have set any kinetic coefficients in the rates equal to one. As a partition we choose $\chi_\alpha = \{1\}$ and $\chi_{\alpha'} = \{2, 3\}$ and in the long time limit the mesostates will thermalize appropriately for any initial state,

$$\begin{pmatrix} \pi_\alpha \\ \pi_{\alpha'} \end{pmatrix} = \lim_{t \to \infty} \begin{pmatrix} p_\alpha(t) \\ p_{\alpha'}(t) \end{pmatrix} = \frac{1}{e^{-\epsilon} + 2} \begin{pmatrix} 1 \\ 1 + e^{-\epsilon} \end{pmatrix}, \quad (94)$$

i.e., the rate matrix $W(\lambda_{t^*})$ is *ergodic*.

As emphasized above, the conditional microstates need not be in equilibrium initially and we parametrize them by $p_{2|\alpha'}(t^*) = \gamma$, $p_{3|\alpha'}(t^*) = 1 - \gamma$ ($\gamma \in [0, 1]$). In principle, it is possible to analytically compute the generator (20) for the ME at the mesolevel, but we refrain from showing the resulting very long expression. Instead, we focus on Fig. 4. It clearly shows that the IFP of the dynamics is given by Eq. (94) only if we choose $p_{2|\alpha'}(t^*) = \pi_{2|\alpha'}(\lambda_{t^*})$ and $p_{3|\alpha'}(t^*) = \pi_{3|\alpha'}(\lambda_{t^*})$ [implying $\gamma = \gamma_{\text{eq}} \equiv e^\epsilon/(1 + e^\epsilon)$], i.e., if the microstates are conditionally equilibrated in agreement with Theorem 4. We have also checked that the time-dependent rates of the generator (20) are always positive for this example (not shown here for brevity) and, hence, the dynamics is 1-Markovian.

This example proves that ergodicity does not imply that $\pi_\alpha(\lambda_t)$ is the IFP of the reduced dynamics, as claimed in

Ref. [71] for arbitrary non-Markovian dynamics. Even 1-Markovianity together with ergodicity is not sufficient to ensure this statement.

B. Markovianity without timescale separation

We give a simple example of a physically relevant and lumpable Markov process although TSS does not apply. For this purpose, consider the following rate matrix:

$$W = \begin{pmatrix} -2\gamma_{\text{in}} & \gamma_{\text{out}} & \gamma_{\text{out}} \\ \gamma_{\text{in}} & -\gamma_{\text{out}} - \bar{\gamma}_{\text{flip}} & \gamma_{\text{flip}} \\ \gamma_{\text{in}} & \bar{\gamma}_{\text{flip}} & -\gamma_{\text{out}} - \gamma_{\text{flip}} \end{pmatrix} \quad (95)$$

describing the time evolution of a probability vector $\mathbf{p}(t) = [p_0(t), p_\uparrow(t), p_\downarrow(t)]$. This ME describes a quantum dot in the ultrastrong Coulomb blockade regime coupled to a metallic lead taking the spin degree of freedom into account. Then, $p_{0/\uparrow/\downarrow}(t)$ are the probabilities to find the dot at time t in a state with zero electrons, an electron with spin up, or an electron with spin down, respectively. If the metallic lead has a finite magnetization, the rates for hopping in (γ_{in}) and out (γ_{out}) of the quantum depend on the spin, which can be derived from first principles [72] and has interesting thermodynamic applications [73]. But, if the lead has zero magnetization as considered here, the dynamics of the spin degree of freedom does not matter. Hence, if we consider the partition $\chi_0 = \{0\}$ and $\chi_1 = \{\uparrow, \downarrow\}$, it is not hard to deduce that

$$\frac{\partial}{\partial t} \begin{pmatrix} p_0(t) \\ p_1(t) \end{pmatrix} = \begin{pmatrix} -2\gamma_{\text{in}} & \gamma_{\text{out}} \\ 2\gamma_{\text{in}} & -\gamma_{\text{out}} \end{pmatrix} \begin{pmatrix} p_0(t) \\ p_1(t) \end{pmatrix}, \quad (96)$$

where $p_1(t) = p_\uparrow(t) + p_\downarrow(t)$. Thus, the coarse-grained dynamics is Markovian for all times t and all microinitial conditions $[p_0(0), p_\uparrow(0), p_\downarrow(0)]$ although TSS does not apply. Notice that the IFP of Eq. (96) coincides with the marginalized IFP of Eq. (95) and, hence, we have $\dot{\Sigma}(t) \geq 0$. Moreover, as long as the structure of the rate matrix (95) is preserved, we could have even allowed for arbitrary time dependencies in the rates.

C. Classical Brownian motion

We here present an example which exhibits negative EP rates and link their appearance to the spectral features of the environment. This is done by considering the important class of driven, classical Brownian motion models (also called Caldeira-Leggett or independent oscillator models). The global Hamiltonian with mass-weighted coordinates reads as

$$H(\lambda_t) = \frac{1}{2}[p^2 + \omega^2(\lambda_t)x^2], \quad (97)$$

$$V + H_B = \frac{1}{2} \sum_k \left[p_k^2 + \nu_k^2 \left(x_k - \frac{c_k}{\nu_k^2}x \right)^2 \right], \quad (98)$$

and its study has attracted considerable interest in strong coupling thermodynamics [32,52,54–59]. The Hamiltonian describes a central oscillator with position x and momentum p linearly coupled to a set of bath oscillators with positions x_k and momenta p_k. The frequency of the central oscillator can

be driven and we parametrize it as $\omega(\lambda_t) = \omega_0 + g \sin(\omega_L t)$. Furthermore, c_k and ν_k are the system-bath coupling constants and the frequencies of the bath oscillators. It turns out that all the information about the bath (except of its temperature) can be encoded into a single function known as the spectral density of the bath. It is defined in general as $J(\omega) \equiv \frac{\pi}{2} \sum_k \frac{c_k^2}{\nu_k} \delta(\omega - \nu_k)$ and we parametrize it as

$$J(\omega) = \frac{\lambda_0^2 \gamma \omega}{\left(\omega^2 - \omega_1^2\right)^2 + \gamma^2 \omega^2}. \qquad (99)$$

Here, λ_0 controls the overall coupling strength between the system and the bath 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. Thus, intuitively one expects that a smaller γ corresponds to stronger non-Markovianity although this intuition can be misleading too [74].

The dynamics of the model is exactly described by the generalized Langevin equation (see, e.g., [75])

$$\ddot{x}(t) + \omega_0^2(t)x(t) + \int_0^t ds\, \Gamma(t-s)\dot{x}(s) = \xi(t) \qquad (100)$$

with the friction kernel

$$\Gamma(t) \equiv \int_0^\infty d\omega \frac{2}{\pi \omega} J(\omega) \cos(\omega t) \qquad (101)$$

and the noise $\xi(t)$, which, when averaged over the initial state of the bath, obeys the statistics

$$\langle \xi(t) \rangle_B = 0, \quad \langle \xi(t)\xi(s) \rangle_B = \frac{1}{\beta}\Gamma(t-s). \qquad (102)$$

To compute the thermodynamic quantities introduced in Sec. IV, we need the state of the system $\rho_S(t)$. It can be computed with the method explained in Sec. IV of Ref. [32], which we will not repeat here. Instead, we focus on the explanation of the numerical observations only.

Figure 5 gives illustrative examples of the time evolution of the EP $\Sigma(t) \geqslant 0$ defined in Eq. (71) for various situations. In total, we plot it for four different parameters characterizing the spectral density, always for the same initial condition of the system, but for the case of an undriven (left column) or a driven (right column) process. The parameters are chosen from top to bottom such that the spectral density resembles more and more an Ohmic spectral density $J(\omega) \sim \omega$, which usually gives rise to Markovian behavior. In fact, this standard intuition is nicely confirmed in Fig. 5 by observing that negative EP rates are much larger and much more common at the top. The plot at bottom indeed corresponds to the Markovian limit in which the bath is conditionally equilibrated throughout (this is similar to the limit of TSS treated in Sec. III C, see also Ref. [32] for additional details). It is worthwhile to repeat that a negative EP rate in the left column of Fig. 5 indicates non-Markovian behavior in a rigorous sense, whereas for the right column this is only true in a weaker sense, but it unambiguously shows that the bath cannot be adiabatically eliminated.
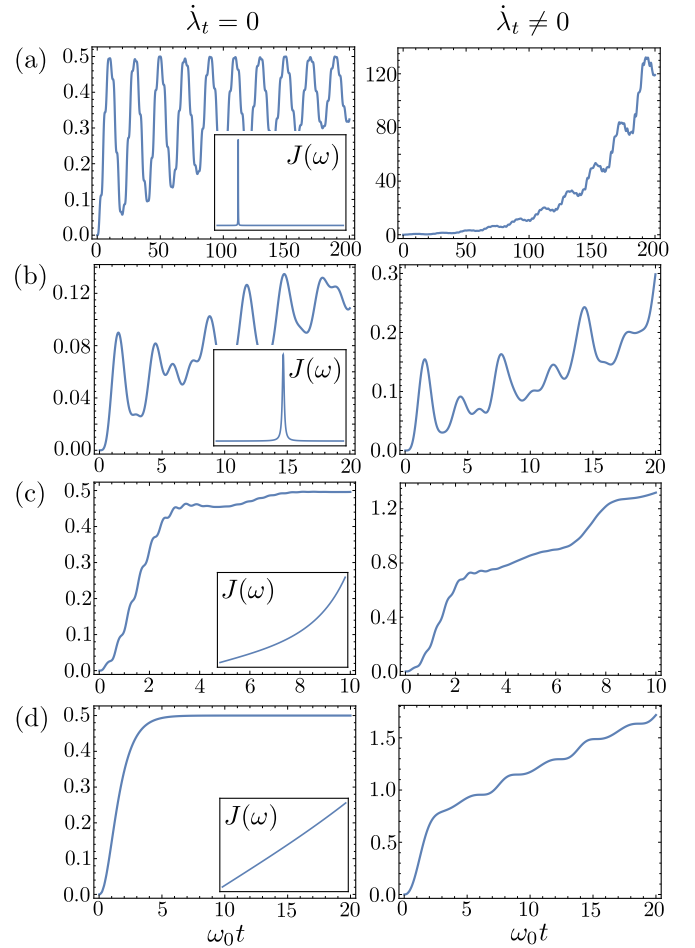

FIG. 5. Plot of the dimensionless entropy production $\Sigma(t)$ ($k_B \equiv 1$) over the dimensionless time $\omega_0 t$ for different parameters. For the driving we chose $g = 0$ and $0.3\omega_0$ for the left or right column, respectively, and $\omega_L = \omega_0$. We changed the shape of the spectral density $J(\omega)$ in each row, which is depicted for $\omega \in [0, 6\omega_0]$ as a small inset (note that the vertical scaling is different in each inset). Specifically, the parameters $(\lambda_0, \gamma, \omega_1)$ are $(0.316\omega_0, 0.01, 1)\omega_0$ (a), $(3.16\omega_0, 0.1, 3.16)\omega_0$ (b), $(100\omega_0, 1, 10)\omega_0$ (c), $(500\omega_0, 10, 31.6)\omega_0$ (d). The system was prepared according to Eq. (12) with 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 $C_{xp_x}(0) = 0$. Note that this specific choice corresponds to equilibrated covariances, but the mean values are out of equilibrium. The general features of the plot, however, do not change too much for different nonequilibrium initial states. Finally, we set $\omega_0 = 1$ and $\beta = 1$. See also Ref. [32] for details of the computation.

D. Quantum dynamics under the initial product state assumption

We have shown in Sec. V that the definition (24) of the EP rate for classical systems does not properly generalize to the quantum case. Part of the problem could be that we started from an initially correlated state, which complicates the treatment of the dynamics of the quantum system significantly. Therefore, one often resorts to the initial product state assumption $\hat{\rho}_{\text{tot}}(0) = \hat{\rho}_S(0) \otimes \hat{\rho}_B$, where $\hat{\rho}_S(0)$ is arbitrary and $\hat{\rho}_B$ fixed (usually taken to be the Gibbs state of the bath) [4,7,8,39,53]. It is then interesting to ask which general

statements connecting Markovianity, the notion of an IFP and EP rates can be made in this case. The following simple example shows which statements do *not* hold in this case.

A single fermionic mode (such as a quantum dot in the Coulomb blockade regime) tunnel coupled to a bath of free fermions (describing, e.g., a metallic lead) can be modeled by the single resonant level Hamiltonian (assuming spin polarization)

$$\hat{H}_{\text{tot}} = \epsilon_0 \hat{d}^\dagger \hat{d} + \sum_k (t_k \hat{d} \hat{c}_k^\dagger + t_k^* \hat{c}_k \hat{d}^\dagger + \epsilon_k \hat{c}_k^\dagger \hat{c}_k). \quad (103)$$

Here, $\hat{d}^{(\dagger)}$ and $\hat{c}_k^{(\dagger)}$ are fermionic annihilation (creation) operators, ϵ_0 is the real-valued energy of the quantum dot, t_k is a complex tunnel amplitude, and ϵ_k is the real-valued energy of a bath fermion.

To describe the dynamics of the open system, we use the Redfield ME [4,39]

$$\frac{\partial}{\partial t} \hat{\rho}_S(t) = -i[\hat{H}, \hat{\rho}_S(t)]$$

$$- \int_0^t ds \, \text{tr}_B\{[\hat{V}, [\hat{V}(s-t), \hat{\rho}_S(t) \otimes \hat{\pi}_B]]\}. \quad (104)$$

Here, the system and interaction Hamiltonian are $\hat{H} = \epsilon_0 \hat{d}^\dagger \hat{d}$ and $\hat{V} = \sum_k (t_k \hat{d} \hat{c}_k^\dagger + t_k^* \hat{c}_k \hat{d}^\dagger)$. Furthermore, $\hat{V}(t) = e^{i(\hat{H}+\hat{H}_B)t/\hbar} \hat{V} e^{-i(\hat{H}+\hat{H}_B)t/\hbar}$ denotes the interaction picture with $\hat{H}_B = \sum_k \epsilon_k \hat{c}_k^\dagger \hat{c}_k$. We assumed the initial system-bath state to be $\hat{\rho}_S(0) \otimes \hat{\pi}_B$ where $\hat{\rho}_S(0)$ is arbitrary and $\hat{\pi}_B$ the grand-canonical equilibrium state with respect to $\hat{H}_B$ and the particle number operator $\hat{N}_B = \sum_k \hat{c}_k^\dagger \hat{c}_k$. Without loss of generality, we set the chemical potential to zero ($\mu = 0$). The Redfield equation (104) directly results from a perturbative expansion of the exact time-convolutionless ME and it gives accurate results for sufficiently small tunneling amplitudes t_k and a relatively high bath temperature.

Following standard procedures, we rewrite Eq. (104) as

$$\frac{\partial}{\partial t} \hat{\rho}_S(t) = -i\epsilon(t)[\hat{d}^\dagger \hat{d}, \hat{\rho}_S(t)]$$

$$+ \gamma_{\text{out}}(t)\left(\hat{d} \hat{\rho}_S(t) \hat{d}^\dagger - \frac{1}{2}\{\hat{d}^\dagger \hat{d}, \hat{\rho}_S(t)\}\right)$$

$$+ \gamma_{\text{in}}(t)\left(\hat{d}^\dagger \hat{\rho}_S(t) \hat{d} - \frac{1}{2}\{\hat{d} \hat{d}^\dagger, \hat{\rho}_S(t)\}\right), \quad (105)$$

where $\{\cdot, \cdot\}$ denotes the anticommutator and $\epsilon(t) \equiv \epsilon_0 - \Delta_{\text{in}}(t) - \Delta_{\text{out}}(t)$ is a time-dependent renormalized system energy. In detail, we have introduced the quantities

$$\gamma_{\text{in}}(t) \equiv \int_0^t d\tau \int_{-\infty}^{\infty} d\omega \frac{J(\omega)}{\pi} f(\omega) \cos[(\omega - \epsilon_0)\tau], \quad (106)$$

$$\Delta_{\text{in}}(t) \equiv \int_0^t d\tau \int_{-\infty}^{\infty} d\omega \frac{J(\omega)}{2\pi} f(\omega) \sin[(\omega - \epsilon_0)\tau], \quad (107)$$

$$\gamma_{\text{out}}(t) \equiv \int_0^t d\tau \int_{-\infty}^{\infty} d\omega \frac{J(\omega)}{\pi} [1 - f(\omega)] \cos[(\omega - \epsilon_0)\tau], \quad (108)$$

$$\Delta_{\text{out}}(t) \equiv \int_0^t d\tau \int_{-\infty}^{\infty} d\omega \frac{J(\omega)}{2\pi} [1 - f(\omega)] \sin[(\omega - \epsilon_0)\tau], \quad (109)$$
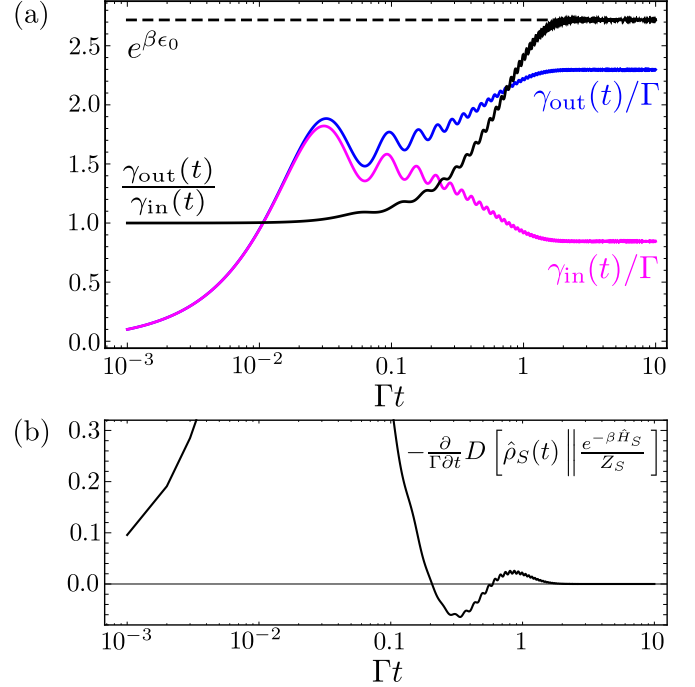

FIG. 6. (a) Plot of the (dimensionless) rates $\gamma_{\text{in}}(t)/\Gamma$ and $\gamma_{\text{out}}(t)/\Gamma$ defined in Eqs. (106) and (108), their ratio $\gamma_{\text{out}}(t)/\gamma_{\text{in}}(t)$, and the expected local detailed balance ratio $e^{\beta\epsilon_0}$ over dimensionless time Γt in logarithmic scale. For the plot we parametrized the bath spectral density as $J(\omega) = \Gamma$ for $\omega/\Gamma \in [-100, +100]$ and zero outside. The dot energy and inverse temperature of the bath are set to $\epsilon_0 = \beta = 1$. (b) For the same parameters we plot an often used candidate for the EP rate over dimensionless time Γt in logarithmic scale for the initial state $p_1(0) = 0.1$, $p_0(0) = 0.9$.

where $f(\omega) \equiv (e^{\beta\omega} + 1)^{-1}$ denotes the Fermi function for $\mu = 0$ and $J(\omega) \equiv 2\pi \sum_k |t_k|^2 \delta(\omega - \epsilon_k)$ is the spectral density of the bath. If there are no initial coherences in the quantum dot present, we can conclude without any further approximation that the full dynamics of the quantum dot is captured by the rate ME

$$\frac{\partial}{\partial t}\begin{pmatrix} p_1(t) \\ p_0(t) \end{pmatrix} = \begin{pmatrix} -\gamma_{\text{out}}(t) & \gamma_{\text{in}}(t) \\ \gamma_{\text{out}}(t) & -\gamma_{\text{in}}(t) \end{pmatrix}\begin{pmatrix} p_1(t) \\ p_0(t) \end{pmatrix}, \quad (110)$$

where $p_1(t)$ [$p_0(t)$] describes the probability to find the dot in the filled (empty) state at time t.

We now investigate the IFP of the dynamics. In Fig. 6 (top) we plot the time evolution of the rates $\gamma_{\text{in}}(t)$ and $\gamma_{\text{out}}(t)$ as well as their ratio. We see that for long times they become stationary and their ratio fulfills local detailed balance (33), which implies that the steady state is a Gibbs state and, hence, the system properly thermalizes. However, for short times, the ratio does not fulfill local detailed balance and, hence, the IFP is not the Gibbs state. Furthermore, as the rates are positive all the time, the dynamics is clearly 1-Markovian. This proves that a 1-Markovian time evolution, which yields the correct long-time equilibrium state, can nevertheless have a time-dependent IFP, even if the underlying Hamiltonian is time independent. This clearly shows that 1-Markovian evolution does not imply a time-invariant IFP as claimed in
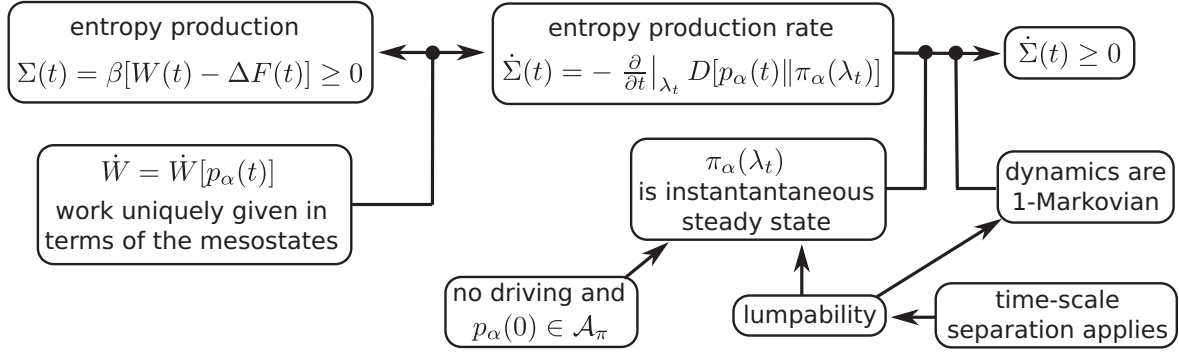
FIG. 7. Overview of the results from Secs. II–IV (the notation is chosen as in Secs. II and III, but the findings are identical to Sec. IV). The arrows indicate implications in a mathematical sense. Some implications depend on certain conditions, which are marked by a line attached with a circle to the respective arrow.

the literature [see, e.g., below Eq. (47) in Ref. [39] or Eq. (9) in Ref. [76]].

In addition, Fig. 6 (bottom) also shows the time evolution of

$$\dot{\sigma}(t) \equiv -\frac{\partial}{\partial t} D[\hat{\rho}_S(t) \| e^{-\beta \hat{H}_S}/Z_S]. \tag{111}$$

In the weak coupling limit, it is tempting to identify $\dot{\sigma}(t)$ as the EP rate because the global equilibrium state can be approximated by $\hat{\pi}_{SB} \approx e^{-\beta \hat{H}_S}/Z_S \otimes e^{-\beta \hat{H}_B}/Z_B$. However, one should be cautious here as this is not an exact result and the initial product state assumption does not fit into the description used in Secs. IV and V. The transient dynamics is indeed dominated by the buildup of system-bath correlations and an exact treatment needs to take them into account [53]. Therefore, outside the specific limit of the Born-Markov secular master equation, where $\dot{\sigma}(t)$ can be related to the actual EP rate [18,20], the quantity $\dot{\sigma}(t)$ lacks a clear connection to a consistent thermodynamic framework. In addition, Fig. 6 clearly demonstrates that $\dot{\sigma}(t) < 0$ is possible although the dynamics is 1-Markovian. For these reasons, the claimed connections between a negative "entropy production" rate $\dot{\sigma}(t)$ and non-Markovianity in Refs. [77–80] require a careful reassessment.

VII. SUMMARY AND OUTLOOK

A. Summary

A large part of this paper was devoted to study the instantaneous thermodynamics at the rate level for an arbitrary classical system coupled to a single heat bath. Quite remarkably, the definition of the EP rate (2) for a weakly coupled Markovian system can be carried over to the strong coupling and non-Markovian situation if we replace the Gibbs state with the correct equilibrium state $\pi_\alpha(\lambda_t)$, described, e.g., by the Hamiltonian of mean force [60]. The EP rate then reads as

$$\dot{\Sigma}(t) \equiv -\frac{\partial}{\partial t}\bigg|_{\lambda_t} D[p_\alpha(t) \| \pi_\alpha(\lambda_t)]. \tag{112}$$

Starting from this definition together with an unambiguous definition for work [Eqs. (43) and (66)], we recovered the previously proposed definitions in Refs. [29,31,32]. Most importantly, we were able to connect the abstract concept of

(non-)Markovianity to the physical observable consequence of having a negative EP rate $\dot{\Sigma}(t) < 0$. We can summarize our finding as follows:

Theorem. If the dynamics is undriven ($\dot{\lambda}_t = 0$), any appearance of $\dot{\Sigma}(t) < 0$ unambiguously reveals that the dynamics is non-Markovian. If the dynamics is driven ($\dot{\lambda}_t \neq 0$), any appearance of $\dot{\Sigma}(t) < 0$ unambiguously reveals that the dynamics is non-Markovian or that $\pi_\alpha(\lambda_t)$ cannot be an IFP of the dynamics. This implies that TSS does not apply.

Especially for the driven case, it was important to study the question of when is the equilibrium state $\pi_\alpha(\lambda_t)$ also an IFP of the dynamics. To the best of our knowledge, this was not yet studied thoroughly. In particular, a 1-Markovian evolution of the system does *not* imply that $\pi_\alpha(\lambda_t)$ is an instantaneous fixed point of the dynamics. This is the reason why a 1-Markovian evolution alone is not sufficient to imply that the entropy production rate is always positive. Figure 7 shows the mathematical implications and equivalences worked out in this paper.

We then left the classical regime and provided a thermodynamic framework for a strongly coupled, driven *quantum* system immersed in an arbitrary heat bath in Sec. V. Inspired by the classical treatment and backed up by equilibrium considerations using the quantum Hamiltonian of mean force [51,66,81], we defined internal energy U, system entropy S, and free energy F [Eqs. (80) to (82)] for a quantum system arbitrarily far from equilibrium. Remarkably, the basic definitions are formally identical to the classical case albeit they were critically debated in Refs. [51,81]. Nevertheless, they ensure that the first and second laws as known from phenomenological nonequilibrium thermodynamics $\Delta U = Q + W$ and $\Sigma = \beta(W - \Delta F) = \Delta S - \beta Q \geqslant 0$ also hold in the quantum regime. Thus, at the integrated level, the quantum nature of the interaction becomes manifest only by realizing that we can treat a smaller class of admissible initially correlated states. At the rate level, however, we showed that the quantum generalization of Eq. (112) does not coincide with the entropy production rate $\dot{\Sigma}(t) = \beta[\dot{W}(t) - d_t F(t)]$. Thus, at present it seems that there is no rigorous connection between negative entropy production rates and non-Markovianity.

To support the latter statement, we also investigated in Sec. VI D what happens for initially decorrelated states if we use the conventional definition of entropy production rate

TABLE I. Current state of the art of strong coupling thermodynamics for a single heat bath. The ♮ symbol indicates only that it is *currently* not known how to establish the corresponding quantum version. Remarks: (a) We here mean that the standard textbook relations between the partition function and internal energy, entropy and free energy are recovered at equilibrium. (b) A work fluctuation theorem of the "Jarzynski-Crooks" type starts with a process in equilibrium and contains the *equilibrium* free energies in the expression. (c) An entropy production (or "integral") fluctuation theorem allows to start in a nonequilibrium state and contains the *nonequilibrium* free energies.

	CM	QM
Consistent with equilibrium thermodynamics[a]	✓	✓
Nonequilibrium first law	✓	✓
Nonequilibrium second law	✓	✓
Recovery of weak coupling limit	✓	✓
Jarzynski-Crooks work fluctuation theorem[b]	✓	✓
Entropy production fluctuation theorem[c]	✓	♮
Arbitrary initial system states	✓	♮
Consistent with TSS	✓	♮
Connection to non-Markovianity	✓	♮

[i.e., the quantum counterpart of Eq. (2)] valid in the limit of the Born-Markov-secular approximation [4,14,18–20]. Unfortunately, outside this limit, this definition does not provide an adequate candidate for an entropy production rate and even for a weakly coupled and 1-Markovian system it can be transiently negative. From the perspective of open quantum system theory, this behavior is caused by the initial buildup of system-environment correlations, which, even in the weak coupling limit, cannot be neglected and need to be taken into account in any formally exact thermodynamic framework [53].

Table I summarizes what is known (and what is not) about the thermodynamic description of a driven system coupled to a single heat bath for the classical (abbreviated CM) and the quantum (QM) cases, respectively.

B. Outlook

After having established a general theoretical description involving a lot of mathematical details, we here take the freedom to be less precise in order to discuss various consequences of our findings and to point out interesting open research avenues.

First of all, the field of strong coupling and non-Markovian thermodynamics is far from being settled and many different approaches have been put forward. Therefore, one might wonder whether the definitions we have used here are the "correct" ones or whether one should not start with a completely different set of definitions. We believe that the definitions we have used possess a certain structural appeal: we could establish a first and second law as known from phenomenological nonequilibrium thermodynamics and in the limit of TSS or at equilibrium, our definitions coincide with established results from the literature. Furthermore, the fact that in the classical case we could give to the appearance of a negative EP rate a

clear dynamical meaning adds further appeal to the definitions used here.

On the other hand, this last point is lost for quantum systems leaving still a larger room of ambiguity there. In this respect, it is also worth to point out that for strongly coupled, non-Markovian systems it was also possible to find definitions which guarantee an always positive EP rate even in presence of multiple heat baths. One possibility is to redefine the system-bath partition [32,56,82–85], which reverses the strategy of Sec. III: instead of looking at the mesostates only when starting from a consistent description in terms of the microstates, one starts with a mesoscopic description and ends up with a consistent description in a larger space, i.e., one effectively finds the microstates from Sec. III. Alternatively and without enlarging the state space, Green's functions techniques can be used for simple models to define an always positive EP rate [86–89] or the polaron transformation can be useful when dealing with particular strong coupling situations [90–94].

Applying our present framework in the context of *multiple* heat baths poses a formidable challenge as it remains unclear what the correct reference state $\pi_\alpha(\lambda_t)$ should be. While it is known how to extend the second law (2) to multiple heat baths if the Born-Markov secular approximation is applied [20], this approximation can be unjustified even at weak coupling [95]. Furthermore, the correct choice of initial state plays a crucial role as it can lead to different thermodynamic definitions; compare, e.g., with the initial product state assumption used in Ref. [53]. At the end, we believe that the most meaningful thermodynamic description will indeed depend on the question which degrees of freedom we can measure and control in an experiment. However, at least at steady state many of the different approaches coincide because the system-bath boundary then usually contributes only a time-independent additive constant to the description.

Within the framework we have used here, we can get also more insights by viewing our findings in light of the recent endeavor to find a meaningful quantifier of non-Markovianity for quantum systems [7,8]. At least for classical, undriven systems it seems reasonable to measure the degree of non-Markovianity via the quantity

$$\mathcal{N} \equiv \max_{p_x(0)\in\mathcal{A}_\pi} \int_{\dot\Sigma(t)<0} |\dot\Sigma(t)| dt \geqslant 0. \tag{113}$$

The larger $\mathcal{N}$, the stronger the system behaves non-Markovian. This quantifier shares structural similarity with the one introduced in Ref. [6] and a nonzero value could be likewise interpreted as information backflow from the bath to the system. Thus, our findings show that due to memory effects, $\dot\Sigma(t)$ loses its property of a Lyapunov function. Of course, $\mathcal{N}$ presents just one out of a multitude of possible non-Markovianity quantifiers [7,8], but it has the outstanding advantage that it is clearly linked to an important and meaningful physical quantity. Its comparison with other measures therefore deserves further attention.

To close this paper, we ask for which problems non-Markovian effects could be beneficial in a thermodynamic sense. This question constitutes in principle a vast field on its own, which we only want to briefly touch. A central benefit of non-Markovian dynamics is that different state

transformations become possible, which are not realizable with a Markovian finite time dynamics.[4] We here want to give a simple example of physical and thermodynamic relevance to illustrate the main point. This example is the erasure of a single bit of information.

Erasing a single bit of information is related to Landauer's famous principle [98] and it is nowadays possible to measure the minuscule thermodynamic changes associated to this transformation [99–105]. Theoretically, the process of erasure is usually modeled with a Markovian two-state system and optimal protocols have been investigated in Refs. [106,107]. Let us now illustrate which benefits non-Markovian dynamics can add. We denote the two states of the bit by "0" and "1" and model the dynamics by the ME

$$\frac{\partial}{\partial t}\begin{pmatrix} p_1(t) \\ p_0(t) \end{pmatrix} = \begin{pmatrix} -\gamma_{01}(t) & \gamma_{10}(t) \\ \gamma_{01}(t) & -\gamma_{10}(t) \end{pmatrix}\begin{pmatrix} p_1(t) \\ p_0(t) \end{pmatrix}. \quad (114)$$

Since we have not made any assumptions about the time-dependent rates $\gamma_{01}(t)$ and $\gamma_{10}(t)$, this model is general and could be obtained directly from Eq. (19). Note that the origin of the time dependence of the rates does not need to come from any driving [cf. Eqs. (19) or (110)]. From $p_0(t) + p_1(t) = 1$ we obtain a linear, inhomogeneous differential equation with time-dependent coefficients for the probability to be in state zero. It reads as $\dot{p}_0(t) = \gamma_{01}(t) - [\gamma_{10}(t) + \gamma_{01}(t)]p_0(t)$ with the formal solution

$$p_0(t) = \exp\left[-\int_0^t ds[\gamma_{10}(s) + \gamma_{01}(s)]\right]p_0(0)$$
$$+ \int_0^t ds \exp\left[-\int_s^t du[\gamma_{10}(u) + \gamma_{01}(u)]\right]\gamma_{01}(s). \quad (115)$$

For definiteness, we choose to erase the bit such that the probability $p_0(t)$ to find the bit in state zero is as large as possible at time t.

Now, as a proof of principle, let us assume that $\gamma_{01}(t) \geqslant 0$ for all times t, but $\gamma_{10}(t)$ can be negative for certain times, which clearly indicates non-Markovian behavior. Furthermore, we denote the fact that $p_0(t)$ depends on the whole history of $\gamma_{10}(t)$ by $p_0(t) = p_0[t; \{\gamma_{10}(t)\}]$. Next, we recall the well-known inequality $\int_0^t ds\, f(s) \leqslant \int_0^t ds |f(s)|$ for any time-dependent function $f(t)$, which implies

$$\exp\left[-\int_0^t ds\, f(s)\right] \geqslant \exp\left[-\int_0^t ds |f(s)|\right]. \quad (116)$$

Because the two terms in Eq. (115) are separately positive, this inequality implies

$$p_0[t; \{\gamma_{10}(t)\}] \geqslant p_0[t; \{|\gamma_{10}(t)|\}] \quad (117)$$

for any initial state and independent of the precise form of the rates. In fact, if for certain times $\gamma_{10}(t) < 0$ we have a strict

inequality: $p_0[t; \{\gamma_{10}(t)\}] > p_0[t; \{|\gamma_{10}(t)|\}]$. This shows that non-Markovian effects can help erase a bit faster in finite time.

To conclude, we believe that our work paves the way for a rigorous understanding of finite-time thermodynamics away from the conventional Markovian assumption. Because our understanding of finite-time processes has drastically improved during the last years [108], exploring their thermodynamic implications opens up a new and exciting research field.

APPENDIX A: WEAK LUMPABILITY

The notion of lumpability required the coarse-grained Markov chain to be Markovian for any initial microstate. One might wonder what can be said about the dynamics if there is at least one initial microstate which leads to a Markov chain at the mesolevel. For this purpose, Kemeny and Snell introduce the concept of weak lumpability (Sec. 6.4 in Ref. [5]):

Definition 4 (Weak lumpability). A Markov chain is weakly lumpable with respect to a partition $\boldsymbol{\chi}$ if there exists at least one initial distribution $p_x(0)$ such that the lumped process is a Markov chain. The TM can then depend on $p_x(0)$.

The fact that the TMs for a weakly lumpable process can depend on the initial microstate $p_x(0)$ is also apparent in Eq. (8). Furthermore, it is again clear that a weakly lumpable process with respect to $p_x(0)$ for a given TM T_τ and partition $\boldsymbol{\chi}$, is also a weakly lumpable process with respect to $p_x(0)$ for all larger times, i.e., for all $T_{n\tau} = (T_\tau)^n$ with $n > 1$ and the same partition $\boldsymbol{\chi}$.

The concept of weak lumpability is especially useful when the underlying Markov chain is regular.[5]

Definition 5 (Regular Markov chain). A Markov chain is called regular if there exists an $n \in \mathbb{N}$ such that all elements of the matrix T_τ^n are strictly positive.

A regular Markov chain ensures that the system reaches its steady state $\boldsymbol{\pi} = \lim_{n\to\infty} T_\tau^n \mathbf{p}(0)$ for any initial distribution $\mathbf{p}(0)$ and, hence, it has a unique steady state. Kemeny and Snell then prove the following [5]:

Theorem 11. Assume that a regular Markov chain with steady state π_x is weakly lumpable with respect to the partition $\boldsymbol{\chi}$ for some initial distribution $p_x(0)$. Then, the Markov chain is also weakly lumpable for the initial distribution π_x with the same transition probabilities, which are determined by

$$G_\tau(\alpha|\beta) = \sum_{x_\alpha, y_\beta} T_\tau(x_\alpha|y_\beta)\pi_{y|\beta}. \quad (A1)$$

---

[4]The question as to whether a given initial state $p_\alpha(0)$ can be transformed into a given final state $p_\alpha(t)$ by a Markovian ME is known as the "embedding problem." For a recent account of this field, see Ref. [96]. The problem was also studied quantum mechanically in Ref. [97].

---

[5]In Sec. VI A, we called this property ergodicity, which is more familiar for a physicist. Here, instead, we follow the terminology of Ref. [5].

Thus, Theorem 11 says that for a regular and weakly lumpable Markov chain, we can always use the conditional steady state $\pi_{x|\alpha}$ to construct the TM at the mesolevel and do not need to use $p_{x|\alpha}(0)$ as in Eq. (8). This is advantageous to say something about the IFP of undriven processes:

Theorem 12. Consider an undriven stochastic process described by the ME (19), i.e., we assume $G_{t,0}^{-1}$ to exist for all admissible initial states $\mathcal{A}(0)$ and all times t. If the stochastic process is weakly lumpable for an underlying regular Markov chain with respect to an admissible initial state $p_x(0) \in \mathcal{A}(0)$, then π_α is a IFP of the stochastic process at the mesolevel.

Proof. Using the insights from Theorem 11, it becomes clear that $\sum_\beta G_{t,0}(\alpha|\beta)\pi_\beta = \pi_\alpha$ for all times t. Together with the invertibility condition we also get Eq. (30) from the main text. These two relations were all we needed to ensure that Eq. (31) holds. ∎

To conclude, as most physically relevant Markov chains are regular, the concept of weak lumpability helps us to deal with initial conditions, where the microstates have not reached a conditional steady state. Together with Theorem 2 this would then imply an always positive EP rate because a weakly lumpable process with respect to an admissible initial state $p_x(0) \in \mathcal{A}(0)$ is also 1-Markovian with respect to that state.

However, a weakly lumpable process still requires the whole hierarchy to fulfill the Markov condition (7) and also on physical grounds we expect that it is a good approximation to assume that the conditional initial microstates are at steady state. If this is the case, then the notion of weak lumpability does not seem to add any further insights into the theory of Secs. III and IV.

APPENDIX B: INSTANTANEOUS FIXED POINTS AND TIME-LOCAL MASTER EQUATION

Formally exact time-local MEs can be derived in different ways. One particular construction was given in Eq. (20), but another possibility is given by the time-convolutionless ME [4,37–39] and see Ref. [35] for yet another way of construction. We will here show that, as long as the inverse of the TM $G_{t,0}$ defined in Eq. (18) exists, the generators all coincide. Hence, the IFP computed with any of those time-local MEs is the same and therefore the IFP is a well-defined concept.

To see this, let us denote by $V^{(1)}(t)$ and $V^{(2)}(t)$ the generators of an exact time-local ME derived in two different ways (we suppress the dependence on λ_t here for simplicity). Because both are assumed to be formally exact for any admissible initial condition, we have

$$\sum_\beta \left[V_{\alpha,\beta}^{(1)}(t) - V_{\alpha,\beta}^{(2)}(t) \right] p_\beta(t) = 0 \tag{B1}$$

for any mesostate $p_\beta(t)$, which is reachable from the class of admissible initial states $\mathcal{A}(0)$. This equation also holds for any linear combination of such states, i.e.,

$$\sum_i \mu_i \sum_\beta \left[V_{\alpha,\beta}^{(1)}(t) - V_{\alpha,\beta}^{(2)}(t) \right] p_\beta^{(i)}(t) = 0 \tag{B2}$$

with $\mu_i \in \mathbb{R}$. We now use that $G_{t,0}$ is invertible for any finite t, which implies in particular that the dimension of the image of $G_{t,0}$ cannot decrease. But, since the class of admissible initial states spans the entire vector space including all probability

distributions $p_\beta(t)$, we can always choose

$$\sum_i \mu_i p_\beta^{(i)}(t) = \delta_{\beta,\beta'} \tag{B3}$$

for any β'. This implies that $V_{\alpha,\beta'}^{(1)}(t) - V_{\alpha,\beta'}^{(2)}(t) = 0$ for any α and β'. Hence, $V^{(1)}(t) = V^{(2)}(t)$.

APPENDIX C: FIXED POINTS OF COARSE-GRAINED HAMILTONIAN DYNAMICS

In this Appendix, we rederive those results from Sec. II, which will be of relevance for Sec. IV. Let us start with an arbitrary Hamiltonian $H(\lambda_t)$ and an arbitrary fixed partition χ. The "master equation" corresponding to this Hamiltonian is the Liouville equation

$$\frac{\partial}{\partial t} \rho(x;t) = \{ H(x;\lambda_t), \rho(x;t) \}, \tag{C1}$$

where $\{\cdot,\cdot\}$ denotes the Poisson bracket and $\rho(x;t)$ is the probability distribution defined on the phase space consisting of the collection of all positions **q** and momenta **p**. For simplicity and analogy with the main text we denote a point in phase space by $x = (\mathbf{q}, \mathbf{p})$. For any partition χ, the mesostates are defined as

$$\rho(\alpha;t) = \int_{\chi_\alpha} dx \, \rho(x;t), \tag{C2}$$

where α can be continuous (e.g., if we trace out a bath) or discrete (e.g., if we lump the motion of a particle in a double well potential into two states "left" and "right"). Furthermore, it turns out to be convenient to denote the dynamical map generated by Eq. (C1) over a finite time interval by $\Phi_{t,0}$, i.e.,

$$\rho(x;t) = \int dx' \Phi_{t,0}(x|x')\rho(x';0), \tag{C3}$$

similar to the time-evolution operator in quantum mechanics.

Clearly, as in Sec. II for a given conditional initial microstate $\rho(x|\alpha;0)$, $\Phi_{t,0}$ induces a map at the mesolevel

$$\rho(\alpha;t) = \int d\beta \, \mathcal{G}_{t,0}(\alpha|\beta)\rho(\beta;0), \tag{C4}$$

$$\mathcal{G}_{t,0}(\alpha|\beta) \equiv \int_{\chi_\alpha} dx \int_{\chi_\beta} dx' \Phi_{t,0}(x|x')\rho(x'|\beta;0). \tag{C5}$$

Using the procedure outlined in Sec. II B or the time-convolutionless ME [4,37–39], we write the time evolution of the mesostate again in terms of a formally exact ME [cf. Eq. (19)]

$$\frac{d}{dt}\rho(\alpha;t) = \mathcal{V}(\lambda_t, t)\rho(\alpha;t). \tag{C6}$$

We note that the Hamiltonian dynamics generated by $\Phi_{t,s}$ $(t \geq s)$ are Markovian. It is therefore possible to straightforwardly extend the definition of lumpability to Hamiltonian dynamics for any propagator $\Phi_{t,s}$. We will here choose a version for the infinitesimal propagator $\Phi_{t+\delta t,t}$ which is most useful in the following.

Definition 6 (Lumpability: continuous version). The dynamics generated by Eq. (C1) is lumpable with respect to the partition χ if for every initial distribution $\rho(x;0)$ the lumped

process is Markovian and the generator $\mathcal{V}(\lambda_t, t)$ in Eq. (C6) does not depend on $\rho(x; 0)$.

We now formulate the analog of Theorem 3:

Theorem 13. If the stochastic process is lumpable as in Definition 6 for some time interval I, then the IFP of the system is given by the marginal global Gibbs state with respect to $H(x; \lambda_t)$ for all $t \in I$, i.e., $\mathcal{V}(\lambda_t, t)\pi(\alpha; \lambda_t) = 0$ with

$$\pi(\alpha; \lambda_t) = \int_{\chi_\alpha} dx \frac{e^{-\beta H(x; \lambda_t)}}{Z(\lambda_t)}. \qquad (C7)$$

Proof. By assumption, the dynamics of the system is generated by

$$\frac{d}{dt}\pi(\alpha; \lambda_t) = \mathcal{V}(\lambda_t, t)\pi(\alpha; \lambda_t)$$

$$= \int_{\chi_\alpha} dx \{H(x; \lambda_t), \pi(\alpha; \lambda_t)\rho(x|\alpha; t)\}, \quad (C8)$$

where $\rho(x|\alpha; t)$ is so far an unknown conditional microstate and where we used that the reduced dynamics from Eq. (C6) is formally exact and, thus, they coincide with the coarse-grained global dynamics.

Next, by assumption of lumpability, we know that $\mathcal{V}(\lambda_t, t)$ is the same *for any initial state*. Let us choose the particular initial state

$$\rho(x; 0) = \Phi_{t,0}^{-1}\pi(x; \lambda_t), \qquad (C9)$$

which is obtained by evolving the Gibbs state $\pi(x; \lambda_t)$ at time t backward in time. Since the global dynamics is Hamiltonian, we remark that the inverse of $\Phi_{t,0}$ exists and maps well-defined probability distribution onto well-defined probability distributions. But, for this choice we clearly have

$$\frac{d}{dt}\pi(\alpha; \lambda_t) = \int_{\chi_\alpha} dx \{H(x; \lambda_t), \pi(x; \lambda_t)\} = 0. \qquad (C10)$$

This implies the theorem. ∎

In principle, of course, we expect the concept of lumpability to be of limited use for Hamiltonian dynamics. However, we can also establish the first part of Theorem 4 for Hamiltonian dynamics:

Theorem 14. Consider an undriven Hamiltonian. If the set of admissible initial states obeys $\mathcal{A}(0) \subset \mathcal{A}_\pi$, then $\pi(\alpha)$ is a IFP of the stochastic process at the mesolevel.

The proof is identical to the first part of the proof of Theorem 4. Furthermore, as in Theorem 5, this also holds for the time-dependent case whenever $\mathcal{A}(t) \subset \mathcal{A}_\pi(\lambda_t)$ or $\pi(x; \lambda_t) \in \mathcal{A}(t)$.

Finally, Theorem 2 then follows analogously by replacing the discrete relative entropy by its differential version and by noting that a proper generalization of Lemma 1 holds also for infinite dimensions [18].

[1] T. L. Hill, *Free Energy Transduction in Biology* (Academic, New York, 1977).

[2] H. Spohn, Kinetic equations from Hamiltonian dynamics: Markovian limits, Rev. Mod. Phys. **52**, 569 (1980).

[3] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. (North-Holland, Amsterdam, 2007).

[4] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2002).

[5] J. G. Kemeny and J. L. Snell, *Finite Markov Chains* (Springer, New York, 1976).

[6] H. P. Breuer, E.-M. Laine, and J. Piilo, Measure for the Degree of Non-Markovian Behavior of Quantum Processes in Open Systems, Phys. Rev. Lett. **103**, 210401 (2009).

[7] A. Rivas, S. F. Huelga, and M. B. Plenio, Quantum non-Markovianity: Characterization, quantification and detection, Rep. Prog. Phys. **77**, 094001 (2014).

[8] H.-P. Breuer, E.-M. Laine, J. Piilo, and B. Vacchini, Colloquium: Non-Markovian dynamics in open quantum systems, Rev. Mod. Phys. **88**, 021002 (2016).

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

[10] D.-Q. Jiang, M. Qian, and M.-P. Qian, *Mathematical Theory of Nonequilibrium Steady States* (Springer, Berlin, 2004).

[11] M. Esposito, U. Harbola, and S. Mukamel, Nonequilibrium fluctuations, fluctuation theorems and counting statistics in quantum systems, Rev. Mod. Phys. **81**, 1665 (2009).

[12] K. Sekimoto, *Stochastic Energetics*, Lecture Notes in Physics (Springer, Berlin, 2010).

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

[14] R. Kosloff, Quantum thermodynamics: A dynamical viewpoint, Entropy **15**, 2100 (2013).

[15] G. Schaller, *Open Quantum Systems Far from Equilibrium*, Lecture Notes in Physics (Springer, Cham, 2014).

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

[17] I. Procaccia and R. D. Levine, Potential work: A statistical mechanical approach for systems in disequilibrium, J. Chem. Phys. **65**, 3357 (1976).

[18] H. Spohn, Entropy production for quantum dynamical semigroups, J. Math. Phys. **19**, 1227 (1978).

[19] G. Lindblad, *Non-Equilibrium Entropy and Irreversibility* (D. Reidel, Dordrecht, Holland, 1983).

[20] H. Spohn and J. L. Lebowitz, Irreversible thermodynamics for quantum systems weakly coupled to thermal reservoirs, Adv. Chem. Phys. **38**, 109 (1979).

[21] B. Altaner, Nonequilibrium thermodynamics and information theory: Basic concepts and relaxing dynamics, J. Phys. A: Math. Theor. **50**, 454001 (2017).

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

[23] D. Andrieux and P. Gaspard, The fluctuation theorem for currents in semi-Markov processes, J. Stat. Mech. (2008) P11007.

[24] M. Esposito and K. Lindenberg, Continuous-time random walk for open systems: Fluctuation theorems and counting statistics, Phys. Rev. E **77**, 051119 (2008).

[25] É. Roldán and J. M. R. Parrondo, Estimating Dissipation from Single Stationary Trajectories, Phys. Rev. Lett. **105**, 150607 (2010).

[26] É. Roldán and J. M. R. Parrondo, Entropy production and Kullback-Leibler divergence between stationary trajectories of discrete systems, Phys. Rev. E **85**, 031129 (2012).

[27] B. Leggio, A. Napoli, H. P. Breuer, and A. Messina, Fluctuation theorems for non-Markovian quantum processes, Phys. Rev. E **87**, 032113 (2013).

[28] B. Bylicka, M. Tukiainen, J. Piilo, D. Chruscinski, and S. Maniscalco, Thermodynamic meaning and power of non-Markovianity, Sci. Rep. **6**, 27989 (2016).

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

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

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

[32] 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).

[33] P. Hänggi and H. Thomas, Time evolution, correlations, and linear response of non-Markov processes, Z. Phys. B **26**, 85 (1977).

[34] A. Rivas, S. F. Huelga, and M. B. Plenio, Entanglement and Non-Markovianity of Quantum Evolutions, Phys. Rev. Lett. **105**, 050403 (2010).

[35] E. Andersson, J. D. Cresser, and M. J. W. Hall, Finding the Kraus decomposition from a master equation and vice versa, J. Mod. Opt. **54**, 1695 (2007).

[36] D. Maldonado-Mundo, P. Öhberg, B. W. Lovett, and E. Andersson, Investigating the generality of time-local master equations, Phys. Rev. A **86**, 042107 (2012).

[37] A. Fulinski and W. J. Kramarczyk, On the exact master equations, Physica (Amsterdam) **39**, 575 (1968).

[38] F. Shibata, Y. Takahashi, and N. Hashitsume, A generalized stochastic Liouville equation. Non-Markovian versus memoryless master equations, J. Stat. Phys. **17**, 171 (1977).

[39] I. de Vega and D. Alonso, Dynamics of non-Markovian open quantum systems, Rev. Mod. Phys. **89**, 015001 (2017).

[40] M. J. W. Hall, J. D. Cresser, L. Li, and E. Andersson, Canonical form of master equations and characterization of non-Markovianity, Phys. Rev. A **89**, 042120 (2014).

[41] G. Nicolis, Transformation properties of entropy production, Phys. Rev. E **83**, 011112 (2011).

[42] M. Esposito and C. Van den Broeck, Three faces of the second law. I. Master equation formulation, Phys. Rev. E **82**, 011143 (2010).

[43] C. Van den Broeck and M. Esposito, Three faces of the second law. II. Fokker-Planck formulation, Phys. Rev. E **82**, 011144 (2010).

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

[45] A. Puglisi, S. Pigolotti, L. Rondoni, and A. Vulpiani, Entropy production and coarse graining in Markov processes, J. Stat. Mech. (2010) P05015.

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

[47] B. Altaner and J. Vollmer, Fluctuation-Preserving Coarse Graining for Biochemical Systems, Phys. Rev. Lett. **108**, 228101 (2012).

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

[49] T. Herpich, J. Thingna, and M. Esposito, Collective Power: Minimal Model for Thermodynamics of Nonequilibrium Phase Transitions, Phys. Rev. X **8**, 031056 (2018).

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

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

[52] E. Aurell, On work and heat in time-dependent strong coupling, Entropy **19**, 595 (2017).

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

[54] E. A. Martinez and J. P. Paz, Dynamics and Thermodynamics of Linear Quantum Open Systems, Phys. Rev. Lett. **110**, 130406 (2013).

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

[56] 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).

[57] N. Freitas and J. P. Paz, Fundamental limits for cooling of linear quantum refrigerators, Phys. Rev. E **95**, 012146 (2017).

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

[59] J.-T. Hsiang, C. H. Chou, Y. Subasi, and B. L. Hu, Quantum thermodynamics from the nonequilibrium dynamics of open systems: Energy, heat capacity, and the third law, Phys. Rev. E **97**, 012135 (2018).

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

[61] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Dissipation: The Phase-Space Perspective, Phys. Rev. Lett. **98**, 080602 (2007).

[62] S. Vaikuntanathan and C. Jarzynski, Dissipation and lag in irreversible processes, Europhys. Lett. **87**, 60005 (2009).

[63] H.-H. Hasegawa, J. Ishikawa, K. Takara, and D. J. Driebe, Generalization of the second law for a nonequilibrium initial state, Phys. Lett. A **374**, 1001 (2010).

[64] K. Takara, H.-H. Hasegawa, and D. J. Driebe, Generalization of the second law for a transition between nonequilibrium states, Phys. Lett. A **375**, 88 (2010).

[65] M. Esposito and C. Van den Broeck, Second law and Landauer principle far from equilibrium, Europhys. Lett. **95**, 40004 (2011).

[66] J.-T. Hsiang and B. L. Hu, Quantum thermodynamics at strong coupling: Operator thermodynamic functions and relations, Entropy **20**, 423 (2018).

[67] A. Uhlmann, Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory, Commun. Math. Phys. **54**, 21 (1977).

[68] M. Ohya and D. Petz, *Quantum Entropy and Its Use* (Springer, Heidelberg, 1993).

[69] R. Uzdin and S. Rahav, Global Passivity in Microscopic Thermodynamics, Phys. Rev. X **8**, 021064 (2018).

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

[71] T. Speck and U. Seifert, The Jarzynski relation, fluctuation theorems and stochastic thermodynamics for non-Markovian processes, J. Stat. Mech. (2007) L09002.

[72] M. Braun, J. König, and J. Martinek, Theory of transport through quantum-dot spin valves in the weak-coupling regime, Phys. Rev. B **70**, 195345 (2004).

[73] P. Strasberg, G. Schaller, T. Brandes, and C. Jarzynski, Second laws for an information driven current through a spin valve, Phys. Rev. E **90**, 062107 (2014).

[74] P. Strasberg and M. Esposito, Response Functions as Quantifiers of Non-Markovianity, Phys. Rev. Lett. **121**, 040601 (2018).

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

[76] G. Thomas, N. Siddharth, S. Banerjee, and S. Ghosh, Thermodynamics of non-Markovian reservoirs and heat engines, Phys. Rev. E **97**, 062108 (2018).

[77] G. Argentieri, F. Benatti, R. Floreanini, and M. Pezzutto, Violations of the second law of thermodynamics by a non-completely positive dynamics, Europhys. Lett. **107**, 50007 (2014).

[78] S. Bhattacharya, A. Misra, C. Mukhopadhyay, and A. K. Pati, Exact master equation for a spin interacting with a spin bath: Non-Markovianity and negative entropy production rate, Phys. Rev. A **95**, 012122 (2017).

[79] S. Marcantoni, S. Alipour, F. Benatti, R. Floreanini, and A. T. Rezakhani, Entropy production and non-Markovian dynamical maps, Sci. Rep. **7**, 12447 (2017).

[80] M. Popovic, B. Vacchini, and S. Campbell, Entropy production and correlations in a controlled non-Markovian setting, Phys. Rev. A **98**, 012130 (2018).

[81] P. Hänggi, G.-L. Ingold, and P. Talkner, Finite quantum dissipation: The challenge of obtaining specific heat, New J. Phys. **10**, 115008 (2008).

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

[83] G. Schaller, J. Cerrillo, G. Engelhardt, and P. Strasberg, Electronic Maxwell demon in the coherent strong-coupling regime, Phys. Rev. B **97**, 195104 (2018).

[84] P. Strasberg, G. Schaller, T. L. Schmidt, and M. Esposito, Fermionic reaction coordinates and their application to an autonomous Maxwell demon in the strong-coupling regime, Phys. Rev. B **97**, 205405 (2018).

[85] S. Restrepo, J. Cerrillo, P. Strasberg, and G. Schaller, From quantum heat engines to laser cooling: Floquet theory beyond the Born-Markov approximation, New J. Phys. **20**, 053063 (2018).

[86] M. Esposito, M. A. Ochoa, and M. Galperin, Quantum Thermodynamics: A Nonequilibrium Greens Function Approach, Phys. Rev. Lett. **114**, 080602 (2015).

[87] A. Bruch, M. Thomas, S. Viola Kusminskiy, F. von Oppen, and A. Nitzan, Quantum thermodynamics of the driven resonant level model, Phys. Rev. B **93**, 115318 (2016).

[88] M. F. Ludovico, M. Moskalets, D. Sánchez, and L. Arrachea, Dynamics of energy transport and entropy production in ac-driven quantum electron systems, Phys. Rev. B **94**, 035436 (2016).

[89] P. Haughian, M. Esposito, and T. L. Schmidt, Quantum thermodynamics of the resonant-level model with driven system-bath coupling, Phys. Rev. B **97**, 085435 (2018).

[90] G. Schaller, T. Krause, T. Brandes, and M. Esposito, Single-electron transistor strongly coupled to vibrations: Counting statistics and fluctuation theorem, New J. Phys. **15**, 033032 (2013).

[91] T. Krause, T. Brandes, M. Esposito, and G. Schaller, Thermodynamics of the polaron master equation at finite bias, J. Chem. Phys. **142**, 134106 (2015).

[92] D. Gelbwaser-Klimovsky and A. Aspuru-Guzik, Strongly coupled quantum heat machines, J. Phys. Chem. Lett. **6**, 3477 (2015).

[93] C. Wang, J. Ren, and J. Cao, Nonequilibrium energy transfer at nanoscale: A unified theory from weak to strong coupling, Sci. Rep. **5**, 11787 (2015).

[94] H. M. Friedman, B. K. Agarwalla, and D. Segal, Quantum energy exchange and refrigeration: A full-counting statistics approach, New J. Phys. **20**, 083026 (2018).

[95] M. T. Mitchison and M. B. Plenio, Non-additive dissipation in open quantum networks out of equilibrium, New J. Phys. **20**, 033005 (2018).

[96] P. Lencastre, F. Raischel, T. Rogers, and P. G. Lind, From empirical data to time-inhomogeneous continuous Markov processes, Phys. Rev. E **93**, 032135 (2016).

[97] M. M. Wolf, J. Eisert, T. S. Cubitt, and J. I. Cirac, Assessing Non-Markovian Quantum Dynamics, Phys. Rev. Lett. **101**, 150402 (2008).

[98] R. Landauer, Irreversibility and heat generation in the computing process, IBM J. Res. Dev. **5**, 183 (1961).

[99] A. O. Orlov, C. S. Lent, C. C. Thorpe, G. P. Boechler, and G. L. Snider, Experimental test of Landauer's principle at the sub-$k_b T$ level, Jpn. J. Appl. Phys. **51**, 06FE10 (2012).

[100] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, Experimental verification of Landauer's principle linking information and thermodynamics, Nature (London) **483**, 187 (2012).

[101] Y. Jun, M. Gavrilov, and J. Bechhoefer, High-Precision Test of Landauer's Principle in a Feedback Trap, Phys. Rev. Lett. **113**, 190601 (2014).

[102] A. Bérut, A. Petrosyan, and S. Ciliberto, Information and thermodynamics: Experimental verification of Landauer's erasure principle, J. Stat. Mech. (2015) P06015.

[103] M. Gavrilov and J. Bechhoefer, Erasure without Work in an Asymmetric Double-Well Potential, Phys. Rev. Lett. **117**, 200601 (2016).

[104] J. Hong, B. Lambson, S. Dhuey, and J. Bokor, Experimental test of Landauer's principle in single-bit operations on nano-magnetic memory bits, Sci. Adv. **2**, e1501492 (2016).

[105] L. L. Yan, T. P. Xiong, K. Rehan, F. Zhou, D. F. Liang, L. Chen, J. Q. Zhang, W. L. Yang, Z. H. Ma, and M. Feng,

Single-Atom Demonstration of the Quantum Landauer Principle, Phys. Rev. Lett. **120**, 210601 (2018).

[106] G. Diana, G. B. Bagci, and M. Esposito, Finite-time erasing of information stored in fermionic bits, Phys. Rev. E **87**, 012111 (2013).

[107] P. R. Zulkowski and M. R. DeWeese, Optimal finite-time erasure of a classical bit, Phys. Rev. E **89**, 052140 (2014).

[108] S. Deffner and S. Campbell, Quantum speed limits: From Heisenbergs uncertainty principle to optimal quantum control, J. Phys. A: Math. Theor. **50**, 453001 (2017).