

# Colloquium: Statistical mechanics and thermodynamics at strong coupling: Quantum and classical

Peter Talkner<sup>✉</sup>

*Institute of Physics, University of Augsburg,  
Universitätsstraße 1, D-86135 Augsburg, Germany  
and Center for Theoretical Physics of Complex Systems, Institute for Basic Science (IBS),  
Daejeon 34113, Republic of Korea*

Peter Hänggi<sup>†</sup>

*Institute of Physics, University of Augsburg,  
Universitätsstraße 1, D-86135 Augsburg, Germany,  
Nanosystems Initiative Munich, Schellingstraße 4, D-80799 München, Germany,  
and Department of Physics, National University of Singapore, Singapore 117546*

 (published 21 October 2020)

The statistical mechanical description of small systems staying in thermal equilibrium with an environment can be achieved by means of the Hamiltonian of mean force. In contrast to the reduced density matrix of an open quantum system, or the reduced phase-space probability density function of a classical open system, the Hamiltonian of mean force not only characterizes the reduced state but also contains full information about the thermodynamics of the considered open system. The resulting thermodynamic potentials all assume the form as the difference of the potentials for the total system and the bare environment in the absence of the system. In contrast to work as a mechanical notion, one faces several problems with the definition of heat, which turns out to be largely ambiguous in the case of strong coupling between system and environment. The general theory of the thermodynamics of open systems, in particular, in view of strong coupling, is reviewed and illustrated it with several examples. The vagueness of heat is discussed in the context of the ambiguities in the definitions of a fluctuating internal energy and other fluctuating thermodynamic potentials.

DOI: [10.1103/RevModPhys.92.041002](https://doi.org/10.1103/RevModPhys.92.041002)

## CONTENTS

I. Introduction	1	2. Work in open classical systems	18
II. Thermodynamics of Large Normal Systems	3	V. Fluctuating Thermodynamic Potentials	18
A. Isolated systems	4	A. Fluctuating internal energy	18
B. Small subsystem of a large closed system	4	B. Fluctuating entropy and free energy	19
III. Subsystem of a Total System at Equilibrium in a Canonical State	5	C. Fluctuating work and heat in open quantum systems	20
A. Hamiltonian of mean force	5	D. Weak coupling	21
B. Thermodynamics	7	VI. Outlook and Conclusions	22
C. Statistical mechanical expressions of the thermodynamic potentials	8	Acknowledgments	22
D. Examples	9	References	23
1. Damped harmonic oscillator	10		
2. Damped free particle	11		
3. Jaynes-Cummings-type model	12		
4. Isotropic $XY$ spin chain	13		
IV. Going into Nonequilibrium	14		
A. Work in thermally isolated systems	14		
1. Classical systems	14		
2. Quantum systems: The two point projective energy measurement scheme (TPPEMS)	15		
B. Work in open systems	17		
1. Work in open quantum systems	17		

## I. INTRODUCTION

Thermodynamics was mainly developed in the 19th century (von Laue, 1950) as a phenomenological theory characterizing equilibrium states of macroscopic bodies and their transformations. In spite of the tremendously large number of microscopic degrees of freedom of a macroscopic system, the number of variables characterizing a thermodynamic equilibrium state is extremely small. For a homogeneous system consisting of a single chemical species, the energy, the mole number, and the volume taken by the system uniquely specify the equilibrium state (Callen, 1985). These variables determine the entropy of the system, which is an extensive function, i.e., a homogeneous function of degree 1, of the variables provided that gravitation, if present at all, can be treated as an external field but does not play a role as an internal interaction. Self-gravitating and other systems with

\*peter.talkner@physik.uni-augsburg.de

†hanggi@physik.uni-augsburg.de

long-range interactions<sup>1</sup> deviate strongly in their behavior from “normal” systems (Kubo, 1965). For example, the entropy of a self-gravitating system is no longer a concave function of the energy everywhere (Thirring, Namhofer, and Posch, 2003).

Statistical mechanics, on the other hand, provides a quantum-mechanical<sup>2</sup> foundation of thermodynamics and yields methods to determine the thermodynamic potentials, such as the entropy, internal, and free energy for specific systems. In principle, these potentials depend on the kind of contact between the considered system and its environment, whether the respective system is thermally isolated or allows an exchange of heat or particles. Technically speaking, ensemble equivalence is tantamount to the concavity of the entropy as a function of the energy. In particular, this property leads to a positive specific heat. Moreover, the transition between different ensembles, say, from the microcanonical to the canonical ensemble, is one to one, given by a Laplace transformation of the density of states. In the thermodynamic limit, this Laplace transformation can be calculated using a steepest descent approximation (Fowler, 1936; Ellis, 1985; Touchette, 2009, 2011) relating the internal energy to the free energy in terms of a Legendre transformation.

For isolated, finite systems the familiar relations valid in the thermodynamic limit of systems with short-range interactions no longer need to be satisfied. Prevailing finite size corrections, depending on the form of the system, such as on the presence of corners, surfaces, and their curvatures, enter the density of states and give rise to deviations from the extensivity of the thermodynamical potentials (Baltes and Hilf, 1976). For a discussion of the proper thermodynamic description of isolated, i.e., microcanonical, systems with a finite and possibly small number of degrees of freedom, we refer to the literature (Schlüter, 1948; Lustig, 1994; Dunkel and Hilbert, 2006; Hilbert, Hänggi, and Dunkel, 2014; Campisi, 2015; Hänggi, Hilbert, and Dunkel, 2016).

A small, weakly interacting part of a large system in a microcanonical state approaches a Maxwell-Boltzmann distribution in the thermodynamic limit of the total system (Khinchin, 1949). Any weakly interacting part of a finite normal system staying in a canonical equilibrium state at some temperature is described by a Maxwell-Boltzmann distribution at the temperature of the total system. It is important to note that under these conditions the only consequence of the presence of the environment is to maintain the equilibrium of the considered small system at a specified temperature. Accordingly, the phase-space distribution of the small system is independent of any properties of the environment other than

---

<sup>1</sup>Even though the bare Coulomb potential of an electrical charge decays in the same way as the gravitational potential, the presence of opposite charges leads to screening such that the thermodynamics of (on the whole neutral) systems consisting of charged constituents does not differ from systems with short-range interactions (Lieb and Lebowitz, 1972).

<sup>2</sup>For simplicity, in the Introduction we refer mostly to classical systems. *Mutatis mutandis*, all statements made here also apply to quantum systems. A distinction between classical and quantum systems is made in later sections, where it becomes necessary to do so.

the temperature and gives rise to the standard thermodynamics of a canonical system independently of its size or form.<sup>3</sup> This universality is lost as soon as one considers the total system as a finite microcanonical system at a specific energy even if the interaction between the proper system and its environment is arbitrarily weak (Campisi, 2007).

In this Colloquium we restrict ourselves to open systems that are possibly strongly interacting parts of larger systems with which they stay in canonical equilibrium at a given temperature. Based on the ergodic hypothesis (Lebowitz and Penrose, 1973), this state of the total system can be realized in different ways, either as an ensemble of microcanonical systems at different energies with an exponential distribution or as a single open system weakly coupling to a superbath at the prescribed temperature. Even though these two scenarios yield identical phase-space distributions of the open system, they are not completely equivalent to each other because a cyclic change of a system parameter within a finite time leads to a change of the ensemble representing the total system at large times while, in the second case, the total system then returns to its initial state due to the presence of the superbath (Talkner and Hänggi, 2016b).

For total systems, which are prepared in either way in a canonical state, the reduced state of the open system may still be written in the form of a Boltzmann distribution at the temperature of the total system with a *Hamiltonian of mean force* (Onsager, 1933; Kirkwood, 1935; Hänggi, Talkner, and Borkovec, 1990; Jarzynski, 2004; Campisi, Talkner, and Hänggi, 2009b) replacing the bare Hamiltonian of the system in the limit of weak coupling. In contrast to the bare system Hamiltonian, the Hamiltonian of mean force depends on temperature and also other parameters determining the microscopic behavior of the environment and its interaction with the open system. While the Hamiltonian of mean force completely specifies the reduced state of the open system, the knowledge of this reduced state in the form of a phase-space probability density function is not sufficient to specify the Hamiltonian of mean force. Therefore, the Hamiltonian of mean force cannot be obtained from a purely system-intrinsic point of view, as discussed in more detail in Sec. III.A. It was argued (Aurell, 2018) that with the embedding techniques of nonlinear dynamics (Badii *et al.*, 1994; Kantz and Schreiber, 1997) the Hamiltonian of the total system can be inferred from observed trajectories of the open system. Given the enormous number of environmental degrees of freedom together with the large amount of data that are required to estimate an unknown Hamiltonian of a system with a few degrees of freedom finding the Hamiltonian of the total system in this way presents, in practice, an impossible task. For quantum systems even a formal procedure corresponding to the embedding technique of classical dynamical systems is missing.

The statistical mechanics of an open system in equilibrium can be specified in the standard way by the Gibbs distribution with the Hamiltonian of mean force, thereby replacing the bare system Hamiltonian. The according partition function of

---

<sup>3</sup>When particles can be exchanged with the environment, a chemical potential is also needed for each particle sort to characterize the environment.

the open system is given by the ratio of the partition functions of the total system and the bare environment (Hänggi and Ingold, 2006). Consequently, all thermodynamic functions of an open system and their derivatives are specified as differences of the respective quantities of the total system and the bare environment. Further, the resulting thermodynamic potentials are *thermodynamically consistent* (Seifert, 2016; Talkner and Hänggi, 2016b) in a sense that is defined later. However, a possible temperature dependence of the Hamiltonian of mean force leads to additional contributions to the statistical mechanical expressions of the internal energy and the thermodynamic entropy, which, in general, does not assume the form of a Shannon or von Neumann entropy (von Neumann, 1955; Wehrl, 1978). As a consequence the thermodynamic entropy of an open system need not be a functional of the reduced state of the open system (Seifert, 2016; Talkner and Hänggi, 2016b); see Eq. (40).

While fluctuations of macroscopic quantities are generally extremely small in macroscopic systems at thermal equilibrium,<sup>4</sup> one may expect that they cannot be neglected in microscopic or mesoscopic systems and hence become an important issue in this context. Because of their time dependence they convey dynamic information about systems in equilibrium. In classical systems one may often identify a set of variables that undergo a Markovian time evolution (Stratonovich, 1963; Hänggi and Thomas, 1982; Gardiner, 1985; Risken, 1989; van Kampen, 2007). For the description of the energetics of such systems, fluctuating heat, work, and internal energy were introduced as “stochastic energetics” (Sekimoto, 2010). With the definition of stochastic entropy (Seifert, 2012) and of further fluctuating thermodynamic potentials (Seifert, 2016; Jarzynski, 2017), a “stochastic thermodynamics” was established recently. Stochastic energetics already suffers from the existence of many random functions, for which the thermal equilibrium averages agree with the correct internal energy of the considered open system. The same flaw also adheres to stochastic thermodynamics because thermodynamic consistency is not sufficient to remove this nonuniqueness. Other restrictions on the hypothetical fluctuating thermodynamic potentials are not known (Talkner and Hänggi, 2016b) for systems other than those weakly coupling to their environments (Talkner, Campisi, and Hänggi, 2009). For quantum systems the work performed on an open system in an individual run of a force protocol can, in principal, be obtained as a fluctuating quantity by means of two projective energy measurements. But even if the experimental techniques to perform projective measurements, such as nondemolition measurements (Braginsky, Vorontsov, and Thorne, 1980; Yang *et al.*, 2020), can be substantially improved, the fact that the work is given by the difference of two often extremely large numbers is seriously limiting the practical accessibility of fluctuating work by means of the two point projective energy measurement scheme (TPPEMS).

For classical systems, the energy difference may be expressed as an integral of the power supplied to the system, and the power can be determined from an observation of the

proper system alone (Liphardt *et al.*, 2001). But even if the work supplied to a classical open system is known, an unambiguous identification of heat, i.e., of the energy that is exchanged within the same process between the system and its environment, is possible only in the weak coupling limit, in which the system-environment interaction is all but neglected. This ambiguity of fluctuating heat is also inherent in the notion of fluctuating energy as the sum of fluctuating work and heat, according to a corresponding formulation of the first law.

The characterization of heat in quantum systems may in principle be based on a TPPEMS of a conveniently defined energy operator of the heat bath, which, once added to the open system internal energy operator, yields the total system Hamiltonian. In processes with a finite interaction time between system and environment, the total transferred heat can be determined by measurements of the bath Hamiltonian before the interaction with the system sets in and after it has ended (Goold, Poschinger, and Modi, 2014). In the case of weak coupling, this environmental energy operator coincides with the bare bath Hamiltonian up to a negligibly small contribution of the system-bath interaction. In all other cases the nonuniqueness of the open system internal energy operator also renders the heat-bath energy operator ambiguous. But also with an arbitrary specification of the bath energy operator, a joint measurement of this operator and the total Hamiltonian cannot be achieved because of their noncommutativity. Hence, it is not possible to simultaneously specify for a quantum process work and heat, not even their averages (Hänggi and Talkner, 2015; Castelvechi, 2017).

## II. THERMODYNAMICS OF LARGE NORMAL SYSTEMS

We first summarize the thermodynamics and statistical mechanics of normal systems (Kubo, 1965), with the goals of recollecting the notions relevant for our main discussion and introducing the notation.

To start with, we refer to a system consisting of a macroscopically large number of microscopic objects like atoms or molecules as normal if in the quantum case the logarithm of the number of states, and in the classical case the logarithm of phase-space volume below a given energy, are homogeneous convex functions of the extensive variables (Kubo, 1965). To satisfy this requirement for a system of classical particles experiencing pairwise interactions in  $d$  spatial dimensions, the interaction potential must be repulsive at short distances and decay with the distance  $r$  faster than  $r^{-d}$  (Ruelle, 1969; Lieb and Lebowitz, 1972). Further, we assume that the dynamics of an autonomous, isolated normal system approaches, after a sufficiently long time, a unique equilibrium state that is independent of the initial state of the system other than its energy. For classical systems this is guaranteed by ergodicity (Khinchin, 1949), for quantum systems the problem of thermalization as such has long been recognized (von Neumann, 1929) but is still under active scrutiny (Deutsch, 1991; Srednicki, 1994; Polkovnikov *et al.*, 2011; Rigol and Srednicki, 2012; Goold *et al.*, 2016; Merali, 2017). In this Colloquium we do not extend this discussion.

<sup>4</sup>Order-parameter fluctuations at second order phase transitions provide a known exception to this rule.

## A. Isolated systems

The dynamics of an isolated system is governed by a Hamiltonian which is the Hamilton function in the case of a classical system and the Hamilton operator for quantum systems. We assume that the gauge of the Hamiltonian is chosen in a way such that it yields the energy of the system even if the parameters  $\lambda$  specifying the Hamiltonian depend on time (Goldstein, 2002). At any fixed set of parameter values the system is supposed to approach an equilibrium state that is completely specified by the energy of the system and hence given by (Bopp, 1953; Münster, 1954)

$$\rho = \omega^{-1}(E, \lambda) \delta(E - H(\lambda)). \quad (1)$$

For a quantum system  $\rho$  presents a density matrix, i.e., a positive operator on the system's Hilbert space with unite trace, and  $\delta$  denotes the Dirac delta function. The normalization of the density matrix is guaranteed by the inverse of the density of states, which is given by

$$\begin{aligned} \omega(E, \lambda) &= \text{Tr} \delta(E - H(\lambda)) \\ &= \sum_n d_n \delta(E - E_n(\lambda)), \end{aligned} \quad (2)$$

where  $E_n(\lambda)$  are the eigenvalues of the Hamiltonian  $H(\lambda)$  and  $d_n$  are the corresponding degrees of degeneracy.<sup>5</sup> Because of the discreteness of the energy spectrum a regularized form of the delta function entering the density matrix must be considered, such as a narrow Gaussian function  $\delta_\epsilon(x) = (2\pi)^{-1/2} \epsilon \exp[-x^2/(2\epsilon)]$  (Talkner, Hänggi, and Morillo, 2008).

For a classical system  $\rho(x)$  presents the probability density function (pdf). It takes the same form as Eq. (1), with  $H(\lambda) = H(\mathbf{x}, \lambda)$  denoting the Hamilton function; the density of states then becomes

$$\omega(E, \lambda) = \int d\mathbf{x} \delta(E - H(\mathbf{x}, \lambda)), \quad (3)$$

with a conveniently defined *dimensionless* infinitesimal phase-space volume  $d\mathbf{x}$ , which allows for indistinguishable particles if necessary, such as  $d\mathbf{x} = d^{3N} p d^{3N} q / N! h^{3N}$  with the Planck constant  $h$  in the case of  $N$  particles in three dimensions, exhibiting no further symmetries. The thermodynamics of a system in a microcanonical state [Eq. (1)] is determined by the microcanonical entropy  $S(E, \lambda)$ , which is given by (Gibbs, 1902; Hertz, 1910a, 1910b; Hilbert, Hänggi, and Dunkel, 2014)

<sup>5</sup>The density of states is not defined for energies from the continuous part of the spectrum. Using the spectral representation of the Hamiltonian one formally obtains with  $\delta(E - H(\lambda)) = \int \delta(E - E') dP(E')$  the expression  $\omega(E, \lambda) = \int \delta(E - E') \text{Tr} dP(E')$ , yielding for an energy belonging to the point spectrum with  $\text{Tr} dP(E_n) = d_n$  the result given in Eq. (2). Here  $P(E)$  denotes the projection operator onto the subspace spanned by all eigenfunctions of  $H$  with energies up to  $E$ . For an energy from the continuous spectrum the trace expression is given by the squared norm of the corresponding eigenfunction, which diverges.

$$S(E, \lambda) = k_B \ln \Omega(E, \lambda), \quad (4)$$

where  $k_B$  is the Boltzmann constant and  $\Omega(E, \lambda)$  specifies the number of states of a quantum system or the phase-space volume of a classical system below the energy  $E$  and hence reads

$$\begin{aligned} \Omega(E, \lambda) &= \int_0^E dE' \omega(E', \lambda) \\ &= \text{Tr} \Theta(E - H(\lambda)) \quad \text{quantum} \\ &= \int d\mathbf{x} \Theta(E - H(\mathbf{x}, \lambda)) \quad \text{classical}. \end{aligned} \quad (5)$$

Here  $\Theta(x)$  denotes the Heaviside function that vanishes for negative arguments  $x$  and yields unity for positive ones. The total differential of the entropy is given by

$$dS = \frac{1}{T} dE + \sum_n \frac{a_n}{T} d\lambda_n, \quad (6)$$

where  $T = (\partial S / \partial E)_\lambda^{-1}$  designates the microcanonical temperature and  $a_i = T(\partial S / \partial \lambda_i)_{E, \lambda_{j \neq i}}$  specifies the response coefficient for a variation of the parameter  $\lambda_i$ . Most notably, the thermodynamic expression for  $a_i$  coincides with its statistical mechanical definition given by the microcanonical average  $a_i = \langle \partial H(\lambda) / \partial \lambda_i \rangle$  (Dunkel and Hilbert, 2014; Hilbert, Hänggi, and Dunkel, 2014). This consistency of statistical mechanics and thermodynamics is guaranteed only for an entropy that depends on the total phase-space volume, as in Eqs. (4) and (5).<sup>6</sup> It agrees<sup>7</sup> for normal systems with the more familiar entropy definition in terms of the density of states, reading

$$S_B(E, \lambda) = k_B \ln[\omega(E, \lambda) \epsilon_B], \quad (7)$$

where  $\epsilon_B$  is an energy scale that must not depend on the values of  $E$  and  $\lambda$ . We note that for quantum systems the phase-space volume entropy [Eq. (4)] is a piecewise constant function of the energy and hence must be smoothed to yield a well-defined temperature and, more generally, to serve as a thermodynamic quantity. The necessary interpolation of the entropy for energies that are different from the eigenvalues of the system Hamiltonian, though, introduces a certain ambiguity.

## B. Small subsystem of a large closed system

The energy of a subsystem fluctuates even if the total system has a fixed energy. The energy fluctuations of an open system follow a Boltzmann distribution provided that the interaction between the considered part and the total system is weak in the sense that the interaction energy is much smaller than the average energy of the subsystem. Additionally, the

<sup>6</sup>The logarithmic dependence of the entropy on the phase-space volume is a consequence of the additivity of the entropy for noninteracting systems.

<sup>7</sup>Up to corrections of the order of  $o(N)$ , with  $N$  denoting the number of microscopic degrees of freedom.

total system must be much larger than the subsystem.<sup>8</sup> The temperature of the Boltzmann distribution of a small system is determined by the microcanonical Boltzmann temperature<sup>9</sup> calculated at the average value of the energy of the large part that provides the environment of the considered open system (Hänggi, Hilbert, and Dunkel, 2016). In the thermodynamic limit, this temperature agrees with the microcanonical temperature of the total system provided that the latter is normal; i.e.,  $T = T_B$ .

Under these conditions, the state of an open system is given by

$$\rho(\beta, \lambda) = Z^{-1}(\beta, \lambda)e^{-\beta H_S(\lambda)}, \quad (8)$$

where  $\beta = (k_B T)^{-1}$  is the inverse temperature and  $H_S(\lambda)$  is the Hamiltonian operator or function of the isolated quantum or classical subsystem, respectively. For a quantum system  $\rho(\beta, \lambda)$  denotes the density matrix and for a classical system the function  $\rho(\mathbf{x}, \beta, \lambda)$  specifies the phase-space probability density function at the phase-space point  $\mathbf{x}$ . The dimensionless partition function  $Z(\beta, \lambda)$  serves for normalization and hence reads<sup>10</sup>

$$Z(\beta, \lambda) = \text{Tr}e^{-\beta H_S(\lambda)} \quad \text{quantum} \quad (9)$$

$$= \int d\mathbf{x} e^{-\beta H_S(\mathbf{x}, \lambda)} \quad \text{classical}. \quad (10)$$

With the knowledge of the partition function the connection between statistical mechanics and thermodynamics is established by

$$F(\beta, \lambda) = -\beta^{-1} \ln Z(\beta, \lambda), \quad (11)$$

which defines the free energy  $F(\beta, \lambda)$ . From this point on, other thermodynamic potentials such as the internal energy  $U(\beta, \lambda)$  and the entropy  $S(\beta, \lambda)$  can be obtained in terms of the textbook relations (Callen, 1985)

$$U(\beta, \lambda) = \left. \frac{\partial(\beta F(\beta, \lambda))}{\partial \beta} \right|_{\lambda}, \quad (12)$$

$$S(\beta, \lambda) = k_B \beta^2 \left. \frac{\partial F(\beta, \lambda)}{\partial \beta} \right|_{\lambda}, \quad (13)$$

which are connected by

<sup>8</sup>When a large system is subdivided into two equally large parts, the energy of each part is approximately Gaussian distributed (Khinchin, 1949).

<sup>9</sup>The Boltzmann temperature  $T_B(E)$  follows from the Boltzmann entropy [Eq. (7)] as  $T_B(E) = [\partial S_B(E, \lambda) / \partial E]_{\lambda}^{-1}$ .

<sup>10</sup>For the partition function to exist, the Gibbsian operator  $e^{-\beta H_S(\lambda)}$  must be an element of the trace class (Schatten, 1950), which is tantamount to the requirement that the Gibbsian operator and consequently also the Hamiltonian have a pure point spectrum but does not contain an absolute or singular continuous part. Moreover, the Hamiltonian must be bounded from below. For classical systems the potential energy must be sufficiently confining to prevent the system from escaping to infinity and also bounded from below.

$$F = U - ST. \quad (14)$$

The joint validity of the three equations (12)–(14) constitutes the *thermodynamic consistency* of the thermodynamic potentials  $F(\beta, \lambda)$ ,  $U(\beta, \lambda)$ , and  $S(\beta, \lambda)$  (Seifert, 2016; Talkner and Hänggi, 2016b). Any two of the three relations imply the remaining one.

In the case of a *weakly coupled open system* in thermal equilibrium, the internal energy and the entropy agree with the standard statistical mechanical expressions for systems in canonical equilibrium, i.e.,

$$U(\beta, \lambda) = \text{Tr} H_S(\lambda) \rho_S(\beta, \lambda), \quad (15)$$

$$S(\beta, \lambda) = -k_B \text{Tr} \rho(\beta, \lambda) \ln \rho(\beta, \lambda), \quad (16)$$

expressing the internal energy  $U(\beta, \lambda)$  as the average value of the bare system Hamiltonian  $H_S$  with respect to the canonical equilibrium state  $\rho(\beta, \lambda)$  specified in Eq. (8) and the entropy  $S(\beta, \lambda)$  as the von Neumann entropy for quantum systems or the Gibbs-Shannon entropy for classical systems. For classical systems, in Eqs. (15) and (16) the trace is to be replaced by the phase-space integral ( $\text{Tr} \rightarrow \int d\mathbf{x}$ ) and the density matrix by the according phase-space pdf.

For a discussion of a small system weakly coupling to a finite bath see Campisi, Talkner, and Hänggi (2009a).

### III. SUBSYSTEM OF A TOTAL SYSTEM AT EQUILIBRIUM IN A CANONICAL STATE

Any canonical equilibrium state can, in principle, be realized in two physically different ways. As described in Sec. II.B, the canonical state may result from the weak contact with another, much larger, system. The time average of the ever changing state of the open system is then given by the canonical state. The other, more formal, way to consider a canonical state is to interpret it as an ensemble of microcanonical states, the energies of which follow a Boltzmann distribution at the given inverse Boltzmann temperature  $\beta$ . While in the first interpretation one considers a single open system, the second scenario consists of many closed systems.

#### A. Hamiltonian of mean force

The Hamiltonian of mean force (Campisi, Talkner, and Hänggi, 2009b) is a fundamental concept in the study of an open system that stays together with its environment in a canonical equilibrium state. It generalizes the notion of the potential of mean force (Onsager, 1933; Kirkwood, 1935; Hänggi, Talkner, and Borkovec, 1990; Roux, 1995) and, in contrast to the latter, it is not restricted to classical situations and can also be assigned to an open quantum system. We first present its definition for quantum systems and later specialize to classical ones.

The starting point is the Hamiltonian  $H_{\text{tot}}$  of the total system, which is composed of contributions describing the bare system and the bare environment  $H_S(\lambda)$  and  $H_B$ , respectively, and an interaction term  $H_{SB}$ . As before, the system Hamiltonian is assumed to depend on a set  $\lambda$  of controllable parameters that must have no influence on the

bare environmental Hamiltonian or the interaction.<sup>11</sup> The total Hamiltonian is therefore given by

$$H_{\text{tot}}(\lambda) = H_S(\lambda) + H_B + H_{SB}. \quad (17)$$

While the canonical thermal equilibrium state of the total system follows with Eq. (8) as

$$\rho_{\text{tot}}(\beta, \lambda) = Z_{\text{tot}}^{-1}(\beta, \lambda) e^{-\beta H_{\text{tot}}(\lambda)}, \quad (18)$$

the state of the open system is determined by the reduced density matrix  $\rho_S(\beta, \lambda)$  and hence becomes

$$\rho_S(\beta, \lambda) = Z_{\text{tot}}^{-1}(\beta, \lambda) \text{Tr}_B e^{-\beta H_{\text{tot}}(\lambda)}, \quad (19)$$

where  $\text{Tr}_B$  denotes the partial trace over the environmental Hilbert space. The reduced density matrix is proportional to the “renormalized Boltzmann factor”  $e^{-\beta H^*(\beta, \lambda)}$  with the Hamiltonian of mean force  $H^*(\beta, \lambda)$ . This renormalized system’s Boltzmann factor results from the Boltzmann factor of the total system by a properly normalized partial trace generating an average over all environmental configurations according to their occurrence in thermal equilibrium, given by

$$e^{-\beta H^*(\beta, \lambda)} = Z_B^{-1}(\beta) \text{Tr}_B e^{-\beta H_{\text{tot}}(\lambda)}. \quad (20)$$

The normalization with the partition function of the bare environment, given by

$$Z_B(\beta) = \text{Tr}_B e^{-\beta H_B}, \quad (21)$$

is uniquely determined by the requirement that for a vanishing system-environment interaction the renormalization yields the result

$$e^{-\beta H^*(\beta, \lambda)} = e^{-\beta H_S(\lambda)} \quad \text{for } H_{SB} = 0 \quad (22)$$

or, equivalently,  $H^*(\beta, \lambda) = H_S(\lambda)$  for the vanishing interaction  $H_{SB}$ . Finally, the reduced state of the open system can be expressed in terms of the Hamiltonian of mean force, yielding

$$\rho_S(\beta, \lambda) = Z_S^{-1}(\beta, \lambda) e^{-\beta H^*(\beta, \lambda)}, \quad (23)$$

with

$$H^*(\beta, \lambda) = -\beta^{-1} \ln \frac{\text{Tr}_B e^{-\beta H_{\text{tot}}(\lambda)}}{\text{Tr}_B e^{-\beta H_B}}. \quad (24)$$

Comparing Eqs. (19) and (23) one finds that the partition function of the open system is given by the ratio of the partition functions of the total system and the bare environment, i.e.,

$$Z_S(\beta, \lambda) = \text{Tr}_S e^{-\beta H^*(\beta, \lambda)} = \frac{Z_{\text{tot}}(\beta, \lambda)}{Z_B(\beta)}, \quad (25)$$

where  $\text{Tr}_S$  denotes the trace over the Hilbert space of the system. The requirement that the renormalization procedure reproduces the bare system Boltzmann factor for a vanishing system-bath interaction was missing in [Gelin and Thoss \(2009\)](#), leading to the erroneous conclusion that the particular form of the partition function as a ratio is arbitrary.

Before we discuss the consequences of this particular structure of the partition function  $Z_S$  for the thermodynamics of an open system, we emphasize the following facts:

- (i) As indicated by the notation and made explicit in specific examples discussed later, the Hamiltonian of mean force depends not only on the parameters  $\lambda$  entering the bare system Hamiltonian but, in general, also on the temperature of the total system, and on all other global parameters as well.
- (ii) Moreover, the structure of the Hamiltonian of mean force depends on the type of environment and its interaction with the open system. Beyond the case of weak coupling, which is considered later, one cannot assume the existence of a generic “thermal environment” the details of which were irrelevant. In general, instead, different environments lead to different Hamiltonians of mean force for the same bare system.
- (iii) Further, we emphasize that the Hamiltonian of mean force does not follow from the reduced state of the open system. From a known reduced density matrix  $\rho_S(\beta, \lambda)$  with the help of Eqs. (23) and (25) one can determine the sum of the Hamiltonian of mean force and the Helmholtz free energy of the open system as

$$H^*(\beta, \lambda) + F_{\text{tot}}(\beta, \lambda) - F_B(\beta) = -\beta^{-1} \ln \rho_S(\beta, \lambda), \quad (26)$$

where we express the logarithms of the partition functions in terms of the respective free energies; see Eq. (11). If there is no additional knowledge about these free energies, the Hamiltonian of mean force remains undetermined.

- (iv) Finally, we note that the partition function  $Z_S(\beta, \lambda)$  defined by Eq. (25) remains finite in the thermodynamic limit of the environment, whereby the number of system degrees of freedom is kept fixed. In this limit the partition functions of the total system as well as that of the environment either diverge or vanish with an increasing number of environmental degrees of freedom,<sup>12</sup> yet their ratio remains finite. In the theory of stochastic energetics ([Sekimoto, 2010](#)) and also in a recent review on stochastic thermodynamics ([Seifert, 2019](#)) coarse-grained, still fluctuating, free energies are introduced without subtraction of the bath contribution. This does not impact free energy differences in an isothermal

<sup>11</sup>This restriction is not necessary here but becomes essential for the definition of work done on an open system by a variation of the parameter  $\lambda$ ; see Sec. IV.B. Other globally acting parameters specifying the Hamiltonian of the total system are generally suppressed so as not to overburden the notation.

<sup>12</sup>We assume here that the total system is normal in the sense of [Kubo \(1965\)](#), as previously explained.

process, but for nonisothermal processes or processes with globally changing parameters it leads to extra contributions to the internal energy and the entropy, as discussed later in more detail.

Equation (24), which specifies the Hamiltonian operator of mean force for a quantum system, can immediately be adapted to classical systems by replacing the Hamiltonian operators with the respective Hamiltonian functions, and the partial trace over the environment with an integral over the environment's phase space, yielding

$$e^{-\beta H^*(\mathbf{x}, \beta, \lambda)} = e^{-\beta H_S(\mathbf{x}, \lambda)} \int d\mathbf{y} e^{-\beta H_{SB}(\mathbf{x}, \mathbf{y})} \rho_B(\mathbf{y}, \beta), \quad (27)$$

where  $\mathbf{x}$  and  $\mathbf{y}$  denote points in the system and the environmental phase space, respectively,  $d\mathbf{y}$  denotes the dimensionless phase-space volume element, and

$$\rho_B(\mathbf{y}, \beta) = Z_B^{-1}(\beta) e^{-\beta H_B(\mathbf{y})} \quad (28)$$

denotes the canonical phase-space pdf of the bare environment, i.e., in the absence of the system. Hence, the classical Hamiltonian of mean force can be expressed as

$$H^*(\mathbf{x}, \beta, \lambda) = H_S(\lambda) - \beta^{-1} \ln \langle e^{-\beta H_{SB}(\mathbf{x}, \mathbf{y})} \rangle_B, \quad (29)$$

with  $\langle \bullet \rangle_B = \int d\mathbf{y} \bullet \rho_B(\mathbf{y}, \beta)$  standing for the average over the canonical state of the bare environment. Note that in the classical case the renormalization is determined by the second term on the right-hand side of Eq. (29), which is independent of the system parameters  $\lambda$ . For the typical case of a system that solely couples the positions of the system  $\mathbf{Q}$  and the environment  $\mathbf{q}$  via the potential  $V_{SB}(\mathbf{Q}, \mathbf{q})$  only the scalar potential of the system is renormalized. The resulting total potential  $V^*(\mathbf{Q})$  is then known as the potential of mean force (Onsager, 1933; Kirkwood, 1935; Hänggi, Talkner, and Borkovec, 1990). It is given by

$$V^*(\mathbf{Q}) = V(\mathbf{Q}) - \beta^{-1} \ln \langle e^{-\beta V_{SB}(\mathbf{Q}, \mathbf{q})} \rangle_B, \quad (30)$$

where  $V(\mathbf{Q})$  denotes the potential of the bare system.

For later use we note the identity

$$\frac{\partial}{\partial \beta} [\beta H^*(\mathbf{x}, \beta, \lambda)] = \langle H_{\text{tot}} | \mathbf{x} \rangle - \langle H_B \rangle_B \quad (31)$$

relating the Hamiltonian of mean force and its inverse temperature derivative to the deviation of the average of the total Hamiltonian conditioned on the state of the system  $\mathbf{x}$  from the average of the bare environmental Hamiltonian (Talkner and Hänggi, 2016b). The average  $\langle \bullet | \mathbf{x} \rangle = \int d\mathbf{y} \bullet w(\mathbf{y} | \mathbf{x})$  is performed with respect to the conditional pdf  $w(\mathbf{y} | \mathbf{x})$  of finding the environment at the phase-space point  $\mathbf{y}$  once the system is at  $\mathbf{x}$ . As such it is given by

$$\begin{aligned} w(\mathbf{y} | \mathbf{x}) &= \frac{\rho_S(\mathbf{x}, \mathbf{y}, \beta, \lambda)}{\rho_S(\mathbf{x}, \beta, \lambda)} \\ &= Z_B^{-1} e^{-\beta(H_{\text{tot}}(\mathbf{x}, \mathbf{y}, \lambda) - H^*(\mathbf{x}, \beta, \lambda))}. \end{aligned} \quad (32)$$

This conditional pdf characterizes the equilibrium preparation class of open classical systems (Grabert, Talkner, and Hänggi, 1977). It plays a key role in the projection operator formulation of the open system's dynamics (Zwanzig, 1961; Grabert, Hänggi, and Talkner, 1980; Grabert, 1982).

In the case of weak coupling (Van Hove, 1957; Davies, 1976) the equilibration of the considered system with its environment is achieved by a vanishingly small interaction strength  $\kappa \rightarrow 0$  at correspondingly late times  $t \rightarrow \infty$  with  $\kappa^2 t$  finite. Accordingly, this weak interaction does not cause a renormalization of the system Hamiltonian, and hence the Hamiltonian of mean force coincides with the Hamiltonian of the bare system. For a further discussion of the weak coupling limit see Sec. V.D.

## B. Thermodynamics

The thermodynamics of an open system being part of a canonical total system follows from its partition function [Eq. (25)] in the standard way starting with the Helmholtz free energy given by

$$F_S(\beta, \lambda) = -\beta^{-1} \ln Z_S(\beta, \lambda). \quad (33)$$

Likewise, one obtains the internal energy and the entropy of the open system by means of the thermodynamic relations (12) and (13). As a consequence of the open system partition function being the ratio of the total system's and the bare bath's functions according to Eq. (25), all thermodynamic potentials and also all response functions result as differences of the respective quantities of the total system and the bare bath. Therefore, they are of the form

$$\Xi_S = \Xi_{\text{tot}} - \Xi_B, \quad (34)$$

where  $\Xi$  stands for any of the thermodynamic potentials. Moreover,  $\Xi$  may stand for the specific heat  $C = \partial U / \partial T$ , and the response to variations of local or global parameters such as the isothermal magnetization  $M_T = \partial F / \partial h |_T$  given by the derivative of the free energy with respect to an external magnetic field  $h$ , isobaric thermal expansion coefficients, as well as for all susceptibilities specified by the second derivatives of the thermodynamic potentials with respect to the relevant local or global parameters of the total system, such as the isothermal magnetic susceptibility  $\chi = \partial^2 F / \partial h^2 |_T$ . Here the indices  $S$ ,  $\text{tot}$ , and  $B$  stand for open system, total system, and bare bath, respectively. This particular form also ensures the validity of the third law of thermodynamics provided that the individual entropies of the total system and the environment vanish when the temperature approaches the absolute zero point (Hänggi and Ingold, 2006). As a difference of two quantities, at finite coupling strength the entropy as well as the specific heat and, at finite coupling, the susceptibilities need *not* comply with their standard positivity properties: The entropy and the specific heat may become negative in certain parameter regions and the susceptibility matrix may also violate positivity. Examples are given later.

Finally, we note that the construction of the thermodynamic potentials of an open system is not restricted to canonical states of the total system but rather can be extended to pressure or grand canonical ensembles with fluctuating volume or particle number, respectively. For example, with the

replacement of the Boltzmann factor  $e^{-\beta H_{\text{tot}}(V)}$  of the total system at fixed volume  $V$  with  $\int dV e^{-\beta[H_{\text{tot}}(V)+pV]}$ , allowing for volume fluctuations controlled by the external pressure  $p$  one obtains in an analogous way for the Hamiltonian of mean force

$$H^*(\beta, p) = -\beta^{-1} \ln \frac{\int dV \text{Tr}_B e^{-\beta[H_{\text{tot}}(V)+pV]}}{\int dV \text{Tr}_B e^{-\beta[H_B(V)+pV]}}, \quad (35)$$

and accordingly for the pressure-dependent partition function of the open system one obtains  $Z_S(\beta, p) = Z_{\text{tot}}(\beta, p)/Z_B(\beta, p)$ , where  $Z_X(\beta, p) = \int dV \text{Tr}_X e^{-\beta(H_X(V)+pV)}$  for  $X = \text{tot}, B$ . A possible dependence on further parameters  $\lambda$  has been suppressed so as not to overburden the notation. All thermodynamic potentials, such as the enthalpy and entropy as well as their derivatives, are obtained from the system Gibbs free energy given by  $G_S(\beta, p) = -\beta^{-1} \ln Z_S(\beta, p)$ . Consequently, they are again determined as the differences of the respective functions of the total system and the bare bath, guaranteeing thermodynamic consistency.<sup>13</sup>

### C. Statistical mechanical expressions of the thermodynamic potentials

The statistical mechanical expression for the free energy of an open system with strong coupling is of the same formal structure as weak coupling, with the Hamiltonian of mean force replacing the bare Hamiltonian. Hence, for the free energy  $F_S(\beta, \lambda)$  one obtains from Eqs. (25) and (33) that

$$F_S(\beta, \lambda) = -\beta^{-1} \ln \text{Tr} e^{-\beta H^*(\beta, \lambda)}. \quad (36)$$

When going from the free energy to the internal energy by means of Eqs. (12), (33), and (36) one finds

$$U_S(\beta, \lambda) = \langle H^*(\beta, \lambda) \rangle_S + \beta \left\langle \frac{\partial H^*(\beta, \lambda)}{\partial \beta} \right\rangle_S, \quad (37)$$

where  $\langle \bullet \rangle_S = \text{Tr}_S \bullet \rho_S(\beta, \lambda)$  denotes the average with respect to the equilibrium system density matrix  $\rho_S(\beta, \lambda) = Z_S^{-1} e^{-\beta H^*(\beta, \lambda)}$  in the quantum case. In the classical case, the trace must be replaced by the respective phase-space integral and the density matrix by the respective phase-space pdf. The first term on the right-hand side corresponds to the standard expression of the internal energy in terms of the equilibrium average of the system Hamiltonian. The second term, which is specific for open systems interacting with their environments at a nonvanishing strength, is a direct consequence of the temperature dependence of the Hamiltonian of mean force. Expressing the internal system energy with Eqs. (17) and (34) as the difference of the internal energies of the total system and the bare bath, one obtains an alternative expression of the

<sup>13</sup>In the so-called bare representation suggested by Jarzynski (2017), the entropy is given by the Shannon entropy of the reduced open system phase-space pdf, and hence the bare representation violates thermodynamic consistency in general beyond the weak coupling regime.

form (Hänggi and Ingold, 2006; Hänggi, Ingold, and Talkner, 2008)

$$U_S(\beta, \lambda) = \langle H_S(\lambda) \rangle_S + \langle H_{SB} \rangle_{\text{tot}} + \langle H_B \rangle_{\text{tot}} - \langle H_B \rangle_B. \quad (38)$$

Here  $\langle \bullet \rangle_{\text{tot}} = \text{Tr}_{SB} \bullet \rho_{\text{tot}}(\beta, \lambda)$  with the trace over the product Hilbert space of the system and the environment  $\text{Tr}_{SB} = \text{Tr}_S \text{Tr}_B$  indicates an average over the canonical state of the total system, and, as previously defined,  $\langle \bullet \rangle_B = \text{Tr}_B \bullet \rho_B(\beta)$  is the average over the canonical state of the bare environment  $\rho_B(\beta) = Z_B^{-1}(\beta) e^{-\beta H_B}$ . The last three terms on the right-hand side of Eq. (38) describe the deviation of the internal energy from the average Hamiltonian of the bare system due to the interaction induced system-environment correlations. They become negligible in the aforementioned weak coupling limit and also in exceptional cases such as for classical open systems with a single degree of freedom in contact with a Caldeira-Leggett-type heat bath (Bogolyubov, 1945; Magalinskii, 1959; Ford, Kac, and Mazur, 1965; Zwanzig, 1973; Caldeira and Leggett, 1983; Hänggi, Talkner, and Borkovec, 1990). Note that in spite of the  $\lambda$  independence of the Hamiltonians  $H_{SB}$  and  $H_B$  their averages with respect to the total equilibrium density matrix  $\rho_{\text{tot}}$  do, in general, depend on  $\lambda$ . Neither these terms nor the internal energy of the bare bath must be neglected, or are only partly taken into account, as one may find in the literature (Sekimoto, 1998; Seifert, 2012; Jarzynski, 2017; Strasberg *et al.*, 2017; Dou *et al.*, 2018; Hsiang *et al.*, 2018).<sup>14</sup>

For the entropy of an open system one finds from Eq. (13) in combination with Eqs. (25) and (33) the expression

$$\begin{aligned} S_S(\beta, \lambda) &= k_B \frac{\partial T \ln Z_S(\beta, \lambda)}{\partial T} \\ &= k_B \ln Z_S - k_B \beta \frac{\partial}{\partial \beta} \ln Z_S \\ &= \mathfrak{S}(\rho_S(\beta, \lambda)) + k_B \beta^2 \left\langle \frac{\partial H^*(\beta, \lambda)}{\partial \beta} \right\rangle_S. \end{aligned} \quad (39)$$

Here  $\mathfrak{S}(\rho)$  denotes the von Neumann entropy of the density matrix  $\rho$ ,  $\mathfrak{S}(\rho) = -k_B \text{Tr} \rho \ln \rho$  for quantum systems (von Neumann, 1955), and for classical systems the continuous Shannon entropy  $\mathfrak{S}(\rho) = -k_B \int d\mathbf{x} \rho(\mathbf{x}) \ln \rho(\mathbf{x})$  of the phase-space pdf with respect to a properly defined dimensionless infinitesimal phase-space volume  $d\mathbf{x}$  (Wehrl, 1978).<sup>15</sup> As for

<sup>14</sup>The argument stated by Seifert (2019) and Strasberg and Esposito (2020), that only differences of the same thermodynamic quantities at the beginning and the end of a thermodynamic process are of relevance, restricts the applicability of the corresponding approach to isothermal processes at constant external parameters. See also the discussion by Talkner and Hänggi (2020).

<sup>15</sup>While the von Neumann entropy remains unchanged under unitary transformations, and likewise the Shannon entropy is invariant under canonical transformations of phase-space variables, the latter is not invariant under general transformations (Marsh, 2013). This is so because  $\rho(\mathbf{x})$  transforms under any transformation  $\mathbf{x} \rightarrow \mathbf{y} = \mathbf{f}(\mathbf{x})$  as a density, i.e.,  $\tilde{\rho}(\mathbf{y}) = \rho(\mathbf{f}^{-1}(\mathbf{y})) / |d\mathbf{y}/d\mathbf{x}|$  and, due to the presence of the Jacobian  $|d\mathbf{y}/d\mathbf{x}|$ , not as a scalar function. This renders the Shannon entropy of a marginal pdf, as for the configuration space or a part of it, an ill defined expression.

the internal energy of open systems outside the weak coupling regime, an additional contribution to the standard von Neumann–Shannon form of entropy exists, in general, due to the temperature dependence of the Hamiltonian of mean force. This guaranties thermodynamic consistency, which means that the free and internal energy and the entropy are related to each other in the standard way expressed in Eq. (14). The presence of the extra term though leads to the fact that the thermodynamic entropy of an open system deviates from the information entropy according to Shannon. Because the Hamiltonian of mean force  $H^*(\beta, \lambda)$  is not uniquely determined by the reduced density matrix  $\rho_S(\beta, \lambda)$ , the thermodynamic entropy cannot be expressed as a functional of the reduced state of the open system alone.

The entropy of an open system can likewise be expressed in terms of the Kullback-Leibler divergence  $S(\rho_{\text{tot}}(\beta, \lambda) \parallel \rho_S(\beta, \lambda) \otimes \rho^B(\beta, \lambda))$  also known as the relative entropy (Wehrl, 1978) between the total state  $\rho_{\text{tot}}(\beta, \lambda)$  and the product state  $\rho_S(\beta, \lambda) \otimes \rho^B(\beta, \lambda)$  and von Neumann entropies of the system and the environment, given by

$$\begin{aligned} S_S(\beta, \lambda) &= -S(\rho_{\text{tot}}(\beta, \lambda) \parallel \rho_S(\beta, \lambda) \otimes \rho^B(\beta, \lambda)) \\ &\quad + \mathfrak{E}(\rho_S(\beta, \lambda)) \\ &\quad + \mathfrak{E}(\rho^B(\beta, \lambda)) - \mathfrak{E}(\rho_B(\beta)). \end{aligned} \quad (40)$$

Here the relative entropy of the density matrices (phase-space pdf's)  $\rho$  and  $\tau$  is defined as  $S(\rho \parallel \tau) = \text{Tr}(\rho(\ln \rho - \ln \tau))$ , and the reduced environmental state  $\rho^B(\beta, \lambda)$  is given by

$$\rho^B(\beta, \lambda) = \text{Tr}_S \rho_{\text{tot}}(\beta, \lambda). \quad (41)$$

Hence, the contribution to the system entropy on the last line of Eq. (40) comes from the difference of the von Neumann entropies of the reduced and the bare environmental state  $\rho^B(\beta, \lambda)$  and  $\rho_B(\beta)$ , respectively; this difference vanishes in the weak coupling limit. In addition, in the latter limit the relative entropy vanishes and, as expected, the entropy of the open system agrees with the von Neumann entropy.

The open system entropy can also be expressed in terms of the conditional entropy of the system, given the state of the environment,  $S(S|B) = \mathfrak{E}(\rho_{\text{tot}}(\beta, \lambda)) - \mathfrak{E}(\rho^B(\beta, \lambda))$  to read

$$S_S(\beta, \lambda) = S(S|B) + \mathfrak{E}(\rho^B(\beta, \lambda)) - \mathfrak{E}(\rho_B(\beta)). \quad (42)$$

Similarly, one may also express the open system entropy  $S_S(\beta, \lambda)$  in terms of the Bayesian sibling  $S(B|S) = \mathfrak{E}(\rho_{\text{tot}}(\beta, \lambda)) - \mathfrak{E}(\rho_S(\beta, \lambda))$  as

$$S_S(\beta, \lambda) = S(B|S) + \mathfrak{E}(\rho_S(\beta, \lambda)) - \mathfrak{E}(\rho_B(\beta)). \quad (43)$$

In passing, we note that for classical systems the conditional entropy of the bath can be written in terms of the conditional pdf  $w(\mathbf{y}|\mathbf{x})$  characterizing the environmental phase-space distribution once the system state  $\mathbf{x}$  of the open system is specified,

$$S(B|S) = -k_B \int d\mathbf{x} d\mathbf{y} \rho_{\text{tot}}(\mathbf{x}, \mathbf{y}, \beta, \lambda) \ln w(\mathbf{y}|\mathbf{x}), \quad (44)$$

where the conditional pdf  $w(\mathbf{y}|\mathbf{x})$  is given by Eq. (32).

While the von Neumann entropy is always positive, none of the equations (34), (39), or (42) guarantee that the entropy of an open system may not take negative values.

Similarly, in view of Eq. (34), the specific heat  $C_S$  and the susceptibilities  $\chi_S$  may assume negative values without indicating any instability of the considered open system (Campisi, Talkner, and Hänggi, 2009c; Ingold, Hänggi, and Talkner, 2009; Campisi, Zueco, and Talkner, 2010). The specific heat can be equally expressed in terms of the partition function as

$$C_S = k_B \beta^2 \frac{\partial^2 \ln Z_S}{\partial \beta^2}. \quad (45)$$

Consequently, in view of the possibility of a negative specific heat, the internal system energy  $U_S = \partial \ln Z_S / \partial \beta$  may decrease with increasing temperature.

#### D. Examples

Before we consider specific examples, we classify those situations in which an open system couples to its environment with a strength beyond the weak coupling limit. As indicated in Fig. 1, we may distinguish among roughly three scenarios. In the first one, sketched in the left panel, a single microscopic object with a few degrees of freedom interacts with its environment; the immediate interaction may be restricted to the close vicinity of this object but may also extend out into more distant parts of the environment. In any case, the weak coupling limit is reached only if, in thermal equilibrium, the interaction energy is negligible compared to the energy of the system degrees of freedom. The second class, sketched in the middle panel of Fig. 1, comprises open systems of mesoscopic, or even macroscopic, size that can be distinguished by their shape and physical properties from the surrounding environment. Here the weak coupling regime sets in when the interaction potential is short range with a characteristic length that is much shorter than the typical linear dimension of the system. Deviations from weak coupling appear in cases of strong short-range interactions and also in relatively weak long-range interactions. The third scenario, pictured in the right panel of Fig. 1, is essential for the theory of solutions (Roux and Simonson, 1999) and can be exemplified by a grain of salt dissolved in a glass of water. As in the second scenario the amount of substance constituting the open system may vary from mesoscopic, say, the size of a cluster of a few hundred atoms, to macroscopic. However, in contrast to the previous case, many of the physical properties including the spatial extension may change fundamentally. In this scenario the weak coupling limit is hardly ever reached. The solution scenario evidently illustrates the influence of the specific properties of the environment on those of the open system: The same solute may behave quite differently in different solvents.

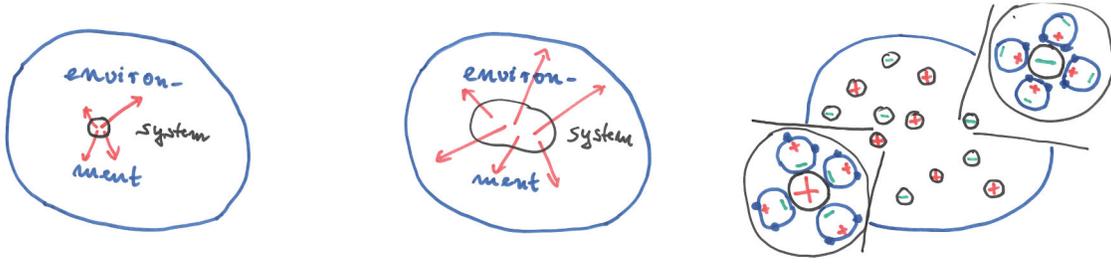


FIG. 1. Different realizations of system-environment coupling. Left panel: a single microscopic entity interacting with the neighborhood of its environment. Middle panel: all parts of a mesoscopic system interacting with at least a part of the environment. This strong coupling situation persists for a macroscopic system if the interaction forces are long-range. Right panel: solute in a solvent such as salt in water. Typically, the dipolar water molecules orient themselves in the vicinity of an ion as a way to shield its charge, as sketched in the magnifications.

### 1. Damped harmonic oscillator

A damped harmonic oscillator in contact with a heat bath at the inverse temperature  $\beta$  can be described in terms of the Zwanzig-Caldeira-Leggett model (Zwanzig, 1973; Caldeira and Leggett, 1983), where an oscillator of mass  $M$ , position  $Q$ , and momentum  $P$  couples to a bath of many other harmonic oscillators mimicking the environment. The Hamiltonian of the total system is given by

$$H_{\text{tot}} = \frac{P^2}{2M} + \frac{1}{2}M\Omega^2 Q^2 + \sum_n \left\{ \frac{p_n^2}{2m_n} + \frac{1}{2}m_n\omega_n^2 \left( q_n - \frac{C_n}{m_n\omega_n^2} Q \right)^2 \right\}, \quad (46)$$

where  $\Omega$  is the frequency of the uncoupled oscillator, while  $m_n$ ,  $\omega_n$ ,  $q_n$ , and  $p_n$  are the mass, frequency, position, and momentum of the  $n$ th bath oscillator. The parameters  $C_n$  determine the coupling strength between the system oscillator and the  $n$ th bath oscillator. For the behavior of the central oscillator, with respect to both its dynamical and equilibrium properties, it is sufficient to specify the so-called memory kernel (Grabert, Schramm, and Ingold, 1988; Weiss, 2008)

$$\gamma(t) = \frac{1}{M} \sum_n \frac{C_n^2}{m_n\omega_n^2} \cos \omega_n t. \quad (47)$$

The partition function  $Z_S(\beta, \Omega)$  can be expressed in terms of the Laplace transform of the memory kernel  $\hat{\gamma}(s) = \int_0^\infty e^{-st} \gamma(t) dt$  as an infinite product of the form (Weiss, 2008)

$$Z_S(\beta, \Omega) = \frac{1}{\beta\hbar\Omega} \prod_{n=1}^{\infty} \frac{\nu_n^2}{\Omega^2 + \nu_n^2 + \nu_n \hat{\gamma}(\nu_n)}, \quad (48)$$

where  $\nu_n = 2\pi n/\hbar\beta$  and  $n = 1, 2, \dots$  denote the Matsubara frequencies. The free energy resulting from Eqs. (33) and (48) agrees with Ford, Lewis, and O'Connell's "remarkable formula" (Ford, Lewis, and O'Connell, 1985) expressing  $F_S(\beta, \Omega)$  as

$$F_S(\beta, \Omega) = \frac{1}{\pi} \int_0^\infty d\omega f(\beta, \omega) \text{Im} \left( \frac{d \ln \chi(\omega)}{d\omega} \right). \quad (49)$$

Here  $f(\beta, \omega) = \beta^{-1} \ln(2 \sinh \beta\hbar\omega/2)$  is the free energy of an isolated harmonic oscillator with frequency  $\omega$  and  $\chi(\omega) = 1/[M(\Omega^2 - \omega^2) - i\omega\hat{\gamma}(i\omega)]$  denotes the susceptibility of the damped oscillator.

For the Drude model, which is specified by  $\gamma(t) = \gamma\omega_D e^{-\omega_D t}$  with the Drude frequency  $\omega_D$  and the static damping constant  $\gamma$ , the partition function of the oscillator can be expressed in closed form (Grabert, Weiss, and Talkner, 1984) as

$$Z_S(\beta, \Omega) = \frac{\beta\hbar\Omega \Gamma(\lambda_1/\nu) \Gamma(\lambda_2/\nu) \Gamma(\lambda_3/\nu)}{4\pi^2 \Gamma(\omega_D/\nu)}, \quad (50)$$

where  $\Gamma(z)$  denotes the gamma function (Abramowitz and Stegun, 1964),  $\lambda_i$ ,  $i = 1, 2, 3$  denotes the solutions of the cubic equation  $\lambda^3 - \omega_D \lambda^2 + (\Omega^2 + \gamma\omega_D)\lambda - \omega_D \Omega^2 = 0$ , and  $\nu = \nu_1$  denotes the fundamental Matsubara frequency. The first moments  $\langle Q \rangle$  and  $\langle P \rangle$  of position and momentum, respectively, vanish; the second moments can be expressed in terms of logarithmic derivatives of the partition function yielding

$$\langle Q^2 \rangle = -\frac{1}{M} \beta \Omega \frac{\partial \ln Z_S}{\partial \Omega}, \quad (51)$$

$$\langle PQ \rangle = \frac{\hbar}{2i}, \quad (52)$$

$$\langle P^2 \rangle = M^2 \Omega^2 \langle Q^2 \rangle - \frac{2M\gamma}{\beta} \frac{\partial \ln Z_S}{\partial \gamma}. \quad (53)$$

Note that the symmetrized position-momentum correlation function  $\langle PQ + QP \rangle$  vanishes because of the time-reversal invariance of the thermal equilibrium state. Because the state of the total system is Gaussian, the reduced density matrix is of Gaussian form and hence is completely determined by its first two moments (Talkner, 1981). The reduced density matrix of the damped oscillator then becomes

$$\rho_S = Z_{\text{eff}}^{-1} e^{-\beta H_{\text{eff}}}, \quad (54)$$

where  $Z_{\text{eff}} = 2 \sinh(\beta \hbar \Omega_{\text{eff}}/2)$  and the effective Hamiltonian is quadratic in position and momentum, given by

$$H_{\text{eff}} = \frac{1}{2M_{\text{eff}}} P^2 + \frac{1}{2} M_{\text{eff}} \Omega_{\text{eff}}^2 Q^2. \quad (55)$$

The renormalized frequency and mass can be expressed as (Grabert, Weiss, and Talkner, 1984)

$$\Omega_{\text{eff}} = \frac{1}{\hbar \beta} \ln \frac{\sqrt{\langle P^2 \rangle \langle Q^2 \rangle} + \hbar/2}{\sqrt{\langle P^2 \rangle \langle Q^2 \rangle} - \hbar/2}, \quad (56)$$

$$M_{\text{eff}} = \frac{1}{\Omega_{\text{eff}}} \sqrt{\frac{\langle P^2 \rangle}{\langle Q^2 \rangle}}. \quad (57)$$

Note that the Hamiltonian of mean force does not coincide with  $H_{\text{eff}}$ , despite other claims (Philbin and Anders, 2016; Miller, 2018), because the resulting normalizing effective partition function  $Z_{\text{eff}} = \text{Tr}_S e^{-\beta H_{\text{eff}}} = [2 \sinh(\beta \hbar \Omega_{\text{eff}}/2)]^{-1}$  does not agree with the open system partition function  $Z_S$  that is given by Eq. (50). The Hamiltonian of mean force is instead given by

$$H^* = H_{\text{eff}} + \beta^{-1} \ln(Z_{\text{eff}}/Z_S). \quad (58)$$

For small temperatures,  $\beta \hbar \Omega_{\text{eff}}$  approaches a finite value, depending on  $\gamma/\Omega$  and  $\omega_D/\Omega$ , with the consequence that the von Neumann entropy  $\mathfrak{S}(\rho_S)/k_B = \beta \hbar \Omega_{\text{eff}}/2 \coth \beta \hbar \Omega_{\text{eff}}/2 - 2 \ln[2 \sinh(\beta \hbar \Omega_{\text{eff}}/2)]$  of the oscillator converges in this limit to a value other than zero, indicating an entanglement between the oscillator and its environment (Hörhammer and Büttner, 2008) in the ground-state wave function of the total system. In contrast to the von Neumann entropy  $\mathfrak{S}(\rho_S)$  the thermodynamic entropy  $S_S(\beta)$  vanishes at low temperatures, in agreement with the third law of thermodynamics (Hänggi and Ingold, 2006; Hänggi, Ingold, and Talkner, 2008; Ingold, Hänggi, and Talkner, 2009).

The classical limit for the damped harmonic oscillator can be performed by letting the dimensionless parameter  $\beta \hbar \Omega$  approach zero, yielding for the partition function  $Z_S$  the classical value  $Z_S^{\text{cl}} = 1/\beta \hbar \Omega$  of a bare harmonic oscillator in a canonical state at the inverse temperature  $\beta$ . Likewise, the Hamiltonian of the mean force approaches in the classical limit the Hamiltonian of the bare oscillator independent of the interaction strength. This must not be seen as specifically classical behavior but instead as a peculiarity of the Zwanzig-Caldeira-Leggett model and related models for systems with a single degree of freedom.<sup>16</sup> The modeling of the environment according to Ullersma (1966) differs from the Zwanzig-Caldeira-Leggett model in the absence of the counterterm

<sup>16</sup>For any total Hamiltonian with a potential function  $U(\{q\}, Q) = V(Q) + F(\{q_n - g_n Q\})$  jointly describing the potential energy of the system and of the environmental degrees of freedom as well as the interaction between system and environment, the integral in Eq. (27) representing the renormalized Boltzmann factor can be performed with the help of the coordinate transformation  $q_n - g_n Q \rightarrow \tilde{q}_n$  for all  $n$ , yielding unity such that Eq. (30) results in  $V^*(Q) = V(Q)$ .

$\sum_k C_k^2/(2m_k \omega_k^2) Q^2$  in the total Hamiltonian. This term then appears with the opposite sign in the classical Hamiltonian of mean force.<sup>17</sup>

Another example displaying a nontrivial Hamiltonian of mean force is given by two particles coupling to the same Zwanzig-Caldeira-Leggett-type environment (Duarte and Caldeira, 2006; Valente and Caldeira, 2010). To be specific, we consider a classical system described by the total Hamiltonian

$$H_{\text{tot}} = H_S + \sum_{n=1}^N \left[ \frac{p_n^2}{2m_n} + \frac{1}{4} m_n \omega_n^2 \sum_{k=1,2} \left( q_n - \frac{2C_{n,k}}{m_n \omega_n^2} Q_k \right)^2 \right], \quad (59)$$

where  $H_S$  is the Hamiltonian of the isolated system with 2 degrees of freedom specified by the coordinates  $Q_1$  and  $Q_2$ , which both couple to the coordinates  $q_n$  of  $N$  harmonic oscillators with coupling strength  $C_{n,k}$ , with  $k = 1, 2$  labeling the system's degrees of freedom. The Hamiltonian of mean force can be calculated by performing Gaussian integrals over the environment degrees of freedom yielding

$$H^* = H_S + \delta V^*, \quad (60)$$

where the potential renormalization is given by

$$\delta V^* = \sum_n \frac{(C_{n,1} Q_1 + C_{n,2} Q_2)^2}{2m_n \omega_n^2}. \quad (61)$$

This potential causes, in general, environment induced forces, both on the center of mass  $Q = (M_1 Q_1 + M_2 Q_2)/(M_1 + M_2)$  and on the relative coordinate  $x = Q_1 - Q_2$ . The force on the center of mass vanishes if the total interaction constants are equal, i.e., for  $\sum_n C_{n,1}/(2m_n \omega_n^2) = \sum_n C_{n,2}/(2m_n \omega_n^2)$ . The force acting on the relative coordinate can be either attractive or repulsive depending on the parameter values. Because of the temperature independence of the potential of mean force, the thermodynamic entropy of the open system  $S_S$ , as specified in Eq. (39), coincides with the Gibbs-Shannon entropy  $\mathfrak{S}(\rho_S)$  of the reduced system pdf  $\rho_S = Z_S^{-1} e^{-\beta H^*}$ . The partition function  $Z_S$  of the open system coincides with the ratio of the total system and the bare environment, in accordance with Eq. (25).

## 2. Damped free particle

In a manner similar to the case of a damped harmonic oscillator, a damped free particle of mass  $M$  can be modeled by means of the Zwanzig-Caldeira-Leggett Hamiltonian (46) by disregarding the parabolic system potential, that is, by setting  $\Omega = 0$  (Grabert, Schramm, and Ingold, 1988). To prevent the particle from escaping to infinity and guarantee the existence of a normalizable thermodynamic equilibrium state, a confining box of large length  $L$  is introduced. In spite of its

<sup>17</sup>For a harmonic oscillator this frequency renormalization may render the system unstable, restricting the choice of the possible environmental parameters. No restrictions of this kind exist for systems with potentials that are more strongly repulsive at infinity than harmonic ones.

seeming simplicity, an exact expression for the partition function of the total system and consequently also for the free damped free particle is not known. Only for sufficiently high temperatures, for which the Gauss sum representing the partition function  $Z_S^0 = \sum_n e^{-\beta E_g n^2}$  of the bare system converges to the respective Gaussian integral,<sup>18</sup> can the partition function  $Z_S$  of the damped particle be approximated by the following expression (Hänggi, Ingold, and Talkner, 2008):

$$Z_S = \frac{L}{\hbar} \left( \frac{2\pi M}{\beta} \right)^{1/2} \frac{\Gamma(1+x_1)\Gamma(1+x_2)}{\Gamma(1+\hbar\beta\omega_D/(2\pi))}, \quad (62)$$

where

$$x_{1,2} = \frac{\hbar\beta\omega_D}{4\pi} (1 \pm \sqrt{1-4\gamma/\omega_D}). \quad (63)$$

Here  $E_g = \pi^2 \hbar^2 / (2ML^2)$  denotes the ground-state energy of the free particle in a box of length  $L$ . Based on the approximate expression (62) of the partition function the specific heat of a damped free particle may be expressed as (Hänggi, Ingold, and Talkner, 2008)

$$C_S/k_B = x_1^2 \psi'(x_1) + x_2^2 \psi'(x_2) - \left( \frac{\hbar\beta\omega_D}{2\pi} \right)^2 \psi' \left( \frac{\hbar\beta\omega_D}{2\pi} \right) - \frac{1}{2}, \quad (64)$$

where  $\psi'(z)$  is the trigamma function (Abramowitz and Stegun, 1964). This expression for the specific heat interpolates between the classical value  $1/2$  (which is reached in the limit of an undamped particle  $\gamma \rightarrow 0$  and in the high temperature limit  $T \rightarrow \infty$ ) and the value  $C = 0$  for  $T \rightarrow 0$  in accordance with the third law of thermodynamics. For strong damping  $\gamma/\omega_D > 1$ , the slope of the specific heat as a function of temperature at  $T = 0$  becomes negative with the consequence that a region of temperatures exists in which the specific heat is negative. This does not indicate any instability of the system but instead the circumstance that, with raising temperature, more energy may be stored in the interaction with the environment than flows into the particle's kinetic energy.

Using Eq. (62) for the partition function in combination with Eq. (13) one finds an entropy that does not vanish for  $T \rightarrow 0$ . This seeming violation of the third law is due to the factorization of the partition function  $Z_S$  in the classical free particle partition function and an environmental term. This approximation disregards the discreteness of the free particle spectrum, thereby leading to a violation of the third law.

An even more complicated behavior of the specific heat was reported by Spreng, Ingold, and Weiss (2013) for other spectral densities than those leading to the Drude model. Coming from positive values at higher temperatures, the specific heat becomes negative for lower temperatures, then positive again for even lower temperatures, but formally fails

<sup>18</sup>A convergence of the Gauss sum to the classical free particle partition function better than 1% is achieved for  $\beta E_g = \beta \pi^2 \hbar^2 / (2ML^2) \lesssim 10^{-4}$ .

to vanish at  $T = 0$ . This apparent violation of the third law is again a consequence of a factorization of partition function analogous to Eq. (62).

### 3. Jaynes-Cummings-type model

As another exactly solvable model we consider a two-level system interacting with an environment made up by a single harmonic oscillator. The Hamiltonian of the total system is

$$H_{\text{tot}} = \frac{\epsilon}{2} \sigma_z + \hbar\omega \left( a^\dagger a + \frac{1}{2} \right) + \kappa \sigma_z \left( a^\dagger a + \frac{1}{2} \right). \quad (65)$$

Here  $\sigma_z$  is the Pauli spin matrix and  $a$  and  $a^\dagger$  are annihilation and creation operators of the oscillator, respectively. The parameters  $\epsilon$ ,  $\omega$ , and  $\kappa$  specify the energy difference of the bare two-level atom, the frequency of the oscillator, and the interaction strength between the two-level atom and the oscillator, respectively. Note that in the Hamiltonian (65) the coupling term commutes with the first two terms describing the free evolution of the oscillator and spin, respectively. Hence, in contrast to the common Jaynes-Cummings model (Jaynes and Cummings, 1963), the interaction is purely dephasing without causing transitions between the eigenstates of the isolated subsystems. For the spectrum of the total Hamiltonian, which is given by  $E_{n,s} = \epsilon s/2 + (\hbar\omega + \kappa s)(n + 1/2)$ ,  $s = \pm 1$ ,  $n = 0, 1, 2, \dots$ , to be bounded from below, the inequality  $\hbar\omega > |\kappa|$  must hold. The equality sign is excluded because otherwise the spectrum contains a point with infinite degeneracy, and therefore the system may not assume a canonical equilibrium.

The partition function of the total system is determined to yield (Campisi, Talkner, and Hänggi, 2009c)

$$Z_{\text{tot}} = q_+ + q_- \quad (66)$$

with the abbreviations

$$q_{\pm} = \frac{e^{-\beta\hbar\omega/2} e^{\mp\beta(\epsilon+\kappa)}}{1 - e^{-\beta(\hbar\omega \pm \kappa)}}. \quad (67)$$

In combination with the partition function of the bare harmonic oscillator  $Z_B = 1/(2 \sinh \beta\hbar\omega/2)$ , the partition function of the open two-level system becomes

$$Z_S = (q_+ + q_-) 2 \sinh(\beta\hbar\omega/2), \quad (68)$$

differing from the partition function of the bare two-level system  $Z_S^0 = 2 \cosh(\beta\epsilon/2)$ .

The Hamiltonian of mean force results in

$$H^* = \frac{\epsilon^*}{2} \sigma_z + \gamma, \quad (69)$$

with the renormalized level distance given by

$$\epsilon^* = \epsilon + \kappa + \frac{2}{\beta} \operatorname{artanh} \left( \frac{e^{-\beta\hbar\omega} \sinh \beta\kappa}{1 - e^{-\beta\hbar\omega} \cosh \beta\kappa} \right), \quad (70)$$

and the energy shift  $\gamma$  given by

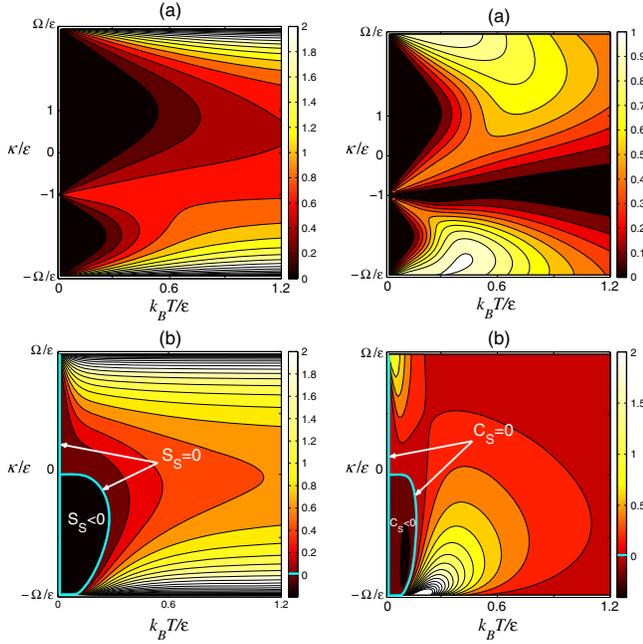


FIG. 2. Role of interaction strength for entropy and specific heat of a two-level system with purely dephasing coupling to an oscillator [Eq. (65)]. The left panels display the entropy and the right panels the specific heat of a two-level atom in contact with a single harmonic oscillator as functions of the dimensionless temperature  $k_B T/\epsilon$  varying on the horizontal axes and of the ratio of the coupling constant and the two-level atom splitting  $\kappa/\epsilon$  within the allowed region  $|\kappa| < \Omega \equiv \hbar\omega$ . The ratio of the oscillator and the atom level splitting is  $\Omega/\epsilon = 3$  in the upper two panels and  $\Omega/\epsilon = 1/3$  in the lower two panels. From Campisi, Talkner, and Hänggi, 2009c.

$$\gamma = \frac{1}{2\beta} \ln \left( \frac{1 - 2e^{-\beta\hbar\omega} \cosh \beta\kappa + e^{-2\beta\hbar\omega}}{(1 - e^{-\beta\hbar\omega})^2} \right). \quad (71)$$

The change of the level spacing  $\Delta = \epsilon^* - \epsilon$  and the energy shift  $\gamma$  vanishes for  $\kappa = 0$  and diverges when the absolute value of  $\kappa$  approaches  $\hbar\omega$ , whereby  $\Delta$  is an odd function and  $\gamma$  is an even function of  $\kappa$ .

The entropy  $S_S = k_B \ln Z_S - k_B \beta \partial \ln Z_S / \partial \beta$  given by Eq. (39) and the specific heat  $C_S = k_B \beta^2 \partial^2 \ln Z_S / \partial \beta^2$  vanish in the limit  $\beta \rightarrow \infty$ , in agreement with the third law of thermodynamics. If the level distance of the harmonic oscillator is less than that of the two-level atom, i.e., if  $\hbar\omega < \epsilon$ , both the entropy and the specific heat become negative at low temperatures for negative coupling constants  $\kappa < 0$ ; see Fig. 2.

#### 4. Isotropic XY spin chain

We consider a linear chain of  $N = N_S + N_B$  spins  $1/2$ , of which the first  $N_S$  spins constitute the system and the remaining  $N_B$  spins constitute the environment sketched in Fig. 3. The chain has free ends and each spin experiences the same magnetic field  $h$  and nearest neighbor interaction of strength  $J$ . It is characterized by the Hamiltonian

$$H_N = \frac{h}{2} \sum_{j=1}^N \sigma_j^z + \frac{J}{2} \sum_{j=1}^{N-1} (\sigma_j^x \sigma_{j+1}^x + \sigma_j^y \sigma_{j+1}^y). \quad (72)$$

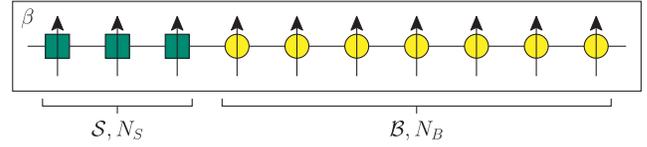


FIG. 3. XY spin chain setup at inverse temperature  $\beta$ . The left  $N_S$  spins depicted as squares constitute the system, while the remaining  $N_B$  spins (circles) compose the environment. From Campisi, Zueco, and Talkner, 2010.

This model is exactly solvable by a Jordan-Wigner transformation (Mikeska and Pesch, 1977) yielding for the partition function

$$Z_N = \text{Tr} e^{-\beta H_N} = e^{-\beta N h / 2} \prod_{k=1}^N (1 + e^{-\beta \lambda_k^{(N)}}), \quad (73)$$

where

$$\lambda_k^{(N)} = h - 2J \cos \frac{\pi k}{N+1}. \quad (74)$$

With the first  $N_S$  spins of this chain as the system and the remaining  $N_B = N - N_S$  spins as the bath, one can recast the total Hamiltonian in the form  $H_{\text{tot}} \equiv H_N = H_S + H_B + H_{SB}$  with

$$H_S = \frac{h}{2} \sum_{j=1}^{N_S} \sigma_j^z - \frac{J}{2} \sum_{j=1}^{N_S-1} (\sigma_j^x \sigma_{j+1}^x + \sigma_j^y \sigma_{j+1}^y), \quad (75)$$

$$H_B = \frac{h}{2} \sum_{j=N_S+1}^N \sigma_j^z - \frac{J}{2} \sum_{j=N_S+1}^{N-1} (\sigma_j^x \sigma_{j+1}^x + \sigma_j^y \sigma_{j+1}^y), \quad (76)$$

$$H_{SB} = -\frac{J}{2} (\sigma_{N_S}^x \sigma_{N_S+1}^x + \sigma_{N_S}^y \sigma_{N_S+1}^y). \quad (77)$$

The partition function of the system part follows as

$$\begin{aligned} Z_S &= \frac{Z_N}{Z_{N-N_S}} \\ &= e^{-\beta N_S h / 2} \frac{\prod_{k=1}^N (1 + e^{-\beta \lambda_k^{(N)}})}{\prod_{k=N_S+1}^N (1 + e^{-\beta \lambda_k^{(N-N_S)}})}. \end{aligned} \quad (78)$$

Figure 4 depicts the specific heat and entropy of open chains of different lengths for two different on-site magnetic fields  $h$  as functions of the temperature. Depending on the strength of the on-site magnetic field and on the length of the open chain, the entropy and the specific heat display regions with negative values. As in the case of negative specific heat observed for a damped free particle at low temperatures, this does not indicate any instability of the system but instead the capacity of the interaction with the environment to effectively store energy when the temperature is rising. For  $T \rightarrow 0$ , both the entropy and the specific heat vanish in accordance with the third law of thermodynamics. The magnetization  $M = \partial F_S / \partial h|_{\beta}$  and the susceptibility  $\chi = \partial M / \partial h|_{\beta}$  following with  $F_S = -\beta^{-1} \ln Z_S$

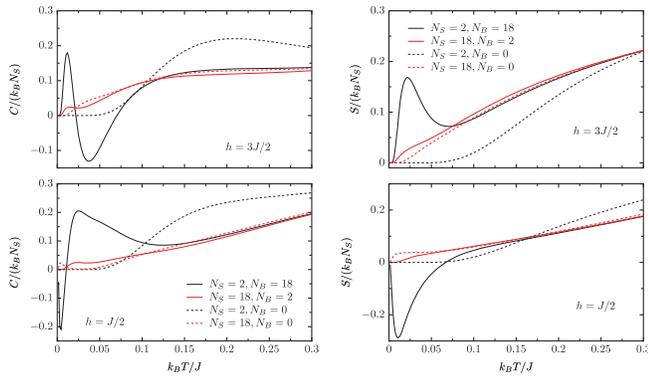


FIG. 4. XY spin chain: specific heat and entropy. The specific heat and the entropy per spin of spin chains are displayed in the left and right panels, respectively, for different magnetic fields  $h = 3J/2$  (upper panels) and  $h = J/2$  (lower panels) and for different combinations of system and bath chain lengths as indicated by different lines as functions of the reduced temperature  $k_B T/J$ . From Campisi, Zueco, and Talkner, 2010.

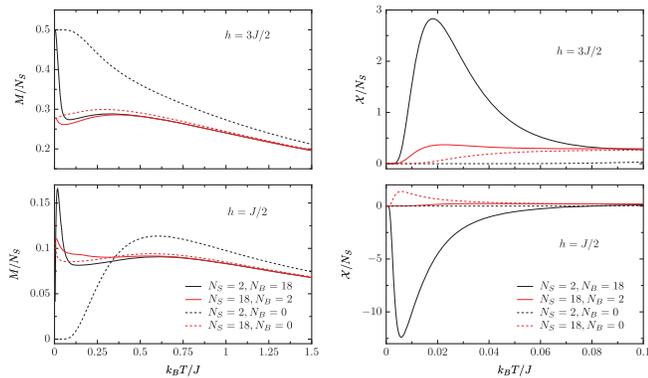


FIG. 5. XY spin chain: magnetization and susceptibility. The magnetization  $M/N_S$  and the susceptibility  $\chi/N_S$  per spin are displayed as functions of the reduced temperature for different chain lengths in the left and right panels, respectively. The upper and lower panels refer to different on-site magnetic fields. From Campisi, Zueco, and Talkner, 2010.

from  $Z_S$  according to Eq. (78) are illustrated for different on-site magnetic fields and chain lengths in Fig. 5.<sup>19</sup> For chains with large interaction  $J > h$ , the susceptibility assumes negative values at low temperatures.

#### IV. GOING INTO NONEQUILIBRIUM

There is a wide variety of circumstances that may drive a system out of thermal equilibrium, either temporarily or permanently (Keizer, 1987; Zwanzig, 2001). Here we restrict ourselves to time-dependent changes of one or several of the system's parameters  $\lambda$ . The energy change involved in such a process is identified with the work done on the system. To properly introduce this notion we first restrict ourselves to

<sup>19</sup>Note that for the magnetization and the susceptibility the subtraction of the bare bath contribution is mandatory because the magnetic field represents a *global* parameter.

isolated systems, then consider open systems in the absence of particle exchange. The pertinent Jarzynski equalities (Jarzynski, 1997a) and Crooks relations (Crooks, 1999) are reviewed in both situations.

#### A. Work in thermally isolated systems

##### 1. Classical systems

The work performed on an isolated system by a time-dependent variation of a system parameter is defined as the resulting change of the systems energy. For a classical system this prescription leads for the work  $w$  performed by the variation of a system parameter  $\lambda(t)$  between  $t=0$  and  $t=\tau$  to

$$w = H(\mathbf{X}(\mathbf{x}, \tau), \lambda(\tau)) - H(\mathbf{x}, \lambda(0)), \quad (79)$$

where  $\mathbf{x}$  denotes a point in the phase space of the system and  $H(\mathbf{x}, \lambda(t))$  denotes the system's Hamiltonian. It is assumed here that this time-dependent Hamiltonian is gauged such that its value coincides with the energy of the system.<sup>20</sup> With  $\mathbf{X}(\mathbf{x}, t)$  denoting the solution of Hamilton's equations of motion starting at  $\mathbf{X}(\mathbf{x}, 0) = \mathbf{x}$  the difference of the Hamiltonians at the final point and at the initial point specifies the difference in energy and hence the work. Here we assume, following Jarzynski (1997a), that the work is determined by the difference of energies resulting from the full Hamiltonians. This work is known as inclusive, or Gibbsian, work. Bochkov and Kuzovlev (1981a, 1981b) used a different definition of work. They assigned as final energy the value of the Hamiltonian at the initial parameter value at the propagated phase-space point, yielding

$$w^{\text{BK}} = H(\mathbf{X}(\mathbf{x}, \tau), \lambda(0)) - H(\mathbf{x}, \lambda(0)). \quad (80)$$

This work definition is also referred to as exclusive work (Jarzynski, 2007). Here we restrict the discussion to the inclusive work defined in Eq. (79). More details on the exclusive work were given by Jarzynski (2007), Campisi, Hänggi, and Talkner (2011), and Campisi, Talkner, and Hänggi (2011). Note that the value of the work depends on both the initial phase-space point  $\mathbf{x}$  and the protocol  $\Lambda = \{\lambda(t) | 0 \leq t \leq \tau\}$ , according to which the parameter  $\lambda(t)$  changes in time. If the initial point  $\mathbf{x}$  is randomly chosen, say, from the canonical equilibrium pdf, the work becomes a random quantity that itself can be characterized by a pdf.

An equivalent formulation of the work defined in Eq. (79) is obtained by rewriting the difference of the Hamiltonians as an integral of the time derivative over the duration of the protocol. Taking into account that the total time derivative of a Hamiltonian as it evolves along a trajectory coincides with its partial derivative, one is led to the following expression for the *classical work*:

$$w = \int_0^\tau dt \frac{\partial H(\mathbf{X}(\mathbf{x}, t), \lambda(t))}{\partial \lambda(t)} \dot{\lambda}(t), \quad (81)$$

<sup>20</sup>For a more exhaustive discussion of the gauge dependence of time-dependent Hamiltonians see Campisi, Hänggi, and Talkner (2011); cf. Sec. III.A therein.

where  $\dot{\lambda}(t)$  denotes the time derivative of  $\lambda(t)$ .<sup>21</sup> The value of the integrand at the time  $t$  gives the power that is supplied to the system by the parameter variation at this instant.

For an initial phase-space pdf  $\rho_0(\mathbf{x})$  the resulting pdf  $p_\Lambda(w)$  of the work is given by (Campisi, Hänggi, and Talkner, 2011)

$$\begin{aligned} p_\Lambda(w) &= \int d\mathbf{x} \delta[w - H(\mathbf{X}(\mathbf{x}, \tau), \lambda(\tau)) + H(\mathbf{x}, \lambda(0))] \rho_0(\mathbf{x}) \\ &= \int d\mathbf{x} \delta\left(w - \int_0^\tau dt \frac{\partial H(\mathbf{X}(\mathbf{x}, t), \lambda(t))}{\partial \lambda(t)} \dot{\lambda}(t)\right) \rho_0(\mathbf{x}). \end{aligned} \quad (82)$$

Focusing on a *canonical* initial state at inverse temperature  $\beta$ , one obtains for the average of the exponentiated negative work per thermal energy  $\beta w$  the Jarzynski equality (Jarzynski, 1997a)

$$\langle e^{-\beta w} \rangle_\Lambda = e^{-\beta \Delta F}, \quad (83)$$

where  $\langle \cdot \rangle_\Lambda = \int dw p_\Lambda(w)$  denotes the average with respect to the work pdf [Eq. (82)], and  $\Delta F = F(\beta, \lambda(\tau)) - F(\beta, \lambda(0))$  is the free energy difference of the canonical equilibrium states at the final and initial parameter values, both at the same temperature. For such isothermal processes  $\Delta F$  corresponds to the maximal work that can be done reversibly by the system. The direct use of the Jarzynski equality, in both numerical and real experiments, as a means of estimating the change of the free energy caused by a variation of a system's parameter is known to often be severely hampered by statistical problems (Lechner and Dellago, 2007; Pohorille, Jarzynski, and Chipot, 2010; Kim *et al.*, 2012; Deng, Jaramillo *et al.*, 2017; Deng, Tan *et al.*, 2017). For classical open systems in contact with a heat bath that are described by a Langevin equation (see Sec. IV.B), the work distribution in the presence of slow forcing can often be well approximated by a Gaussian distribution with a vanishing variance in the quasistatic limit (Hoppenau and Engel, 2013). A similar result holds for open quantum Markovian processes described by a master equation of the Lindblad type (Miller *et al.*, 2019). As a consequence of the Jarzynski equation the irreversible work then vanishes as it does for a quasistatic isothermal macroscopic process. For more general system setups, both the absence and the presence of finite work fluctuations in the quasistatic limit have been reported in the literature (Deng, Tan *et al.*, 2017).

The most remarkable aspect of the Jarzynski equality is that it applies to an arbitrary protocol that is not restricted to be slow. In general, the finally reached state differs from the thermal equilibrium state corresponding to the final parameter value  $\lambda(\tau)$ . Note that for thermally isolated forcing an equilibrium state that might be reached at large times will generally have a temperature differing from the initial one. Only if the system stays in weak contact with a thermal

<sup>21</sup>For a forced parameter  $\lambda(t)$  that couples to the system via a generalized coordinate  $Q(\mathbf{x})$  according to  $H(\mathbf{x}, \lambda(t)) = H_0(\mathbf{x}) - Q(\mathbf{x})\lambda(t)$ , the work expression simplifies to  $w = -\int_0^\tau dt Q(\mathbf{X}(\mathbf{x}, t)) \dot{\lambda}(t)$ . Hence, from an experimentally observed trajectory  $Q(\mathbf{X}(\mathbf{x}, t))$ ,  $t \in [0, \tau]$  the work can be determined.

reservoir having the initial temperature, then the equilibrium state with the system free energy  $F(\beta, \lambda(\tau))$  is approached for an infinitely slow protocol. For fast protocols an equilibration takes place only after a sufficiently long time subsequent to the terminal protocol time  $\tau$ . During the equilibration the so-called irreversible work  $w_{\text{irr}} = w - \Delta F$  is taken by the reservoir. With Jensen's inequality one obtains from the Jarzynski equality that the average of the irreversible work cannot become negative, in agreement with the second law of thermodynamics, i.e.,

$$\langle w \rangle_\Lambda \geq \Delta F, \quad (84)$$

and consequently  $\langle w_{\text{irr}} \rangle_\Lambda \geq 0$ . The average irreversible work is proportional to the Kullback-Leibler divergence of the actually reached final phase-space pdf  $\rho(\tau)$  and the Gibbs state [Eq. (8)]; hence,  $\langle w_{\text{irr}} \rangle_\Lambda = TS[\rho(\tau) \parallel \rho(\beta, \lambda(\tau))]$  (Kawai, Parrondo, and Van den Broeck, 2007).

To any forced process running according to a protocol  $\Lambda = \{\lambda(t) | 0 \leq t \leq \tau\}$  a reverse force protocol  $\bar{\Lambda}(t) = \{\epsilon_\lambda \lambda(\tau - t) | 0 \leq t \leq \tau\}$  can be assigned, where it is assumed that the instant Hamiltonians of the forward and backward processes are related by the time-reversal operation  $H(\mathbf{q}, \mathbf{p}, \lambda) \rightarrow \bar{H}(\mathbf{q}, \mathbf{p}, \lambda) \equiv H(\mathbf{q}, -\mathbf{p}, \epsilon_\lambda \lambda)$ , where  $\lambda$  comprises all parameters on which the Hamiltonian depends, including those that remain fixed during the force protocol. Here  $\epsilon_\lambda$  denotes the parity of the parameter  $\lambda$  under time reversal. For the pair of so-called forward and backward processes, both starting in a canonical equilibrium state at the same inverse temperature  $\beta$  and at the respective parameter values  $\lambda(0)$  and  $\epsilon_\lambda \lambda(\tau)$ , the work pdfs  $p_\Lambda(w)$  and  $p_{\bar{\Lambda}}(w)$  are connected by the Crooks relation (Crooks, 1999)

$$p_\Lambda(w) = e^{-\beta(\Delta F - w)} p_{\bar{\Lambda}}(-w). \quad (85)$$

This implies that the occurrence of work smaller than the free energy change  $\Delta F$  is exponentially small (Jarzynski, 2011), i.e.,

$$P[w < \Delta F - \zeta] \equiv \int_{-\infty}^{\Delta F - \zeta} dw p_\Lambda(w) \leq e^{-\beta \zeta}. \quad (86)$$

This can be understood as a further specification of the second law of thermodynamics (Jarzynski, 2007, 2011).

Multiplying both sides of the Crooks relation [Eq. (85)] by the factor  $e^{-\beta w}$  and integrating over the work one recovers the Jarzynski equality [Eq. (83)].

## 2. Quantum systems: The two point projective energy measurement scheme (TPPEMS)

At first glance the translation of a classical work expression to quantum mechanics might seem obvious by simply replacing the Hamiltonian functions in the equivalent classical work definitions [Eqs. (79) and (81)] with the corresponding operators. This naive approach fails to lead to a proper work operator for various reasons. While in the classical expression [Eq. (79)] the Hamiltonians are evaluated at specific phase-space points that are connected by a trajectory of the Hamiltonian dynamics, the corresponding Hamiltonian

operators are independent of the initial and time-evolved states of the system. For the equivalent classical work expression [Eq. (81)] the phase-space trajectory connecting these states has to be known. Owing to the lack of a classical trajectory in the quantum case, however, this expression cannot be directly converted to a work expression for the quantum world.<sup>22</sup> Further, in classical systems the specification of the energy can, in principle, be performed without any perturbation of the system. For a quantum system, gaining information about the system, from which its energy can be inferred, necessitates the interaction with an auxiliary system such as a measurement apparatus, which, in turn, causes a *backaction* on the considered system. Hence, the specific tools employed to identify energies of a system from which the work is determined constitute a relevant part of any operational definition of quantum work.

For definiteness, we consider the TPPEMS. In this measurement scheme a projective energy measurement is performed on the system in the state  $\rho_0$  immediately before the force protocol  $\Lambda$  starts, i.e., according to our previous convention, at  $t = 0$ , and the second one at  $t = \tau$  immediately after the protocol is finished (Kurchan, 2000; Tasaki, 2000; Talkner, Lutz, and Hänggi, 2007). The joint probability to observe the energies  $E_n(0)$  and  $E_m(\tau)$  is then given by

$$p_\Lambda(m, n) = \text{Tr} P_m(\tau) U_\Lambda P_n(0) \rho_0 P_n(0) U_\Lambda^\dagger. \quad (87)$$

Here  $P_k(t)$  denotes the operator projecting onto the eigenspace of the Hamiltonian  $H(\lambda(t)) = \sum_k E_k(t) P_k(t)$  with eigenenergy  $E_k(t)$ .<sup>23</sup> Because of the first measurement the initial density matrix  $\rho_0$  is projected onto the subspace with  $E_n(0)$  and subsequently propagated by the unitary time-evolution operator

$$U_\Lambda = \mathcal{T} e^{-i \int_0^\tau dt H(\lambda(t)) / \hbar}, \quad (88)$$

where  $\mathcal{T}$  denotes the chronological time-ordering operator. The work pdf  $p_\Lambda(w)$  follows from Eq. (87) as

$$p_\Lambda(w) = \sum_{m,n} \delta(w - E_m(\tau) + E_n(0)) p_\Lambda(m, n). \quad (89)$$

Equivalently, the work statistics can be described by the characteristic function  $G_\Lambda(u) = \int dw e^{i u w} p_\Lambda(w)$ , which assumes the form (Talkner, Lutz, and Hänggi, 2007)

<sup>22</sup>The work calculated along Bohmian trajectories turns out to depend on the particular representation of the initial state in terms of pure states (Sampaio *et al.*, 2018) and hence cannot be considered a measurable quantity.

<sup>23</sup>In this general setting Hamiltonians with spectra containing accumulation points and continuous parts can also be considered. Because of the finite resolution of any measurement apparatus the probability  $p_\Lambda(m, n)$  must then be replaced by the probability  $p_\Lambda(A, B) = \text{Tr} P_A(\tau) U_\Lambda P_B(0) \rho_0 P_B(0) U_\Lambda^\dagger$ , where  $P_C(t)$  projects on all eigenstates with energies  $E(t) \in C$ ,  $C = A, B$  being subsets of the spectrum captured by the two measurements at  $t = 0, \tau$ .

$$G_\Lambda(u) = \text{Tr}(e^{i u H^H(\lambda(\tau))} e^{-i u H(\lambda(0))} \bar{\rho}_0) \quad (90)$$

with the Hamiltonian  $H^H(\lambda(\tau)) = U_\Lambda^\dagger H(\lambda(\tau)) U_\Lambda$  in the Heisenberg picture; further,  $\bar{\rho}_0 = \sum_n P_n(0) \rho_0 P_n(0)$  is the initial density matrix projected onto the energy basis (Talkner, Hänggi, and Morillo, 2008). Here it is worth noting that the projection of the initial state due to the first energy measurement has an impact on the average work  $\langle w \rangle_\Lambda = \text{Tr}[H^H(\lambda(\tau)) - H(\lambda(0))] \bar{\rho}_0$ , which differs from the difference of the average energies at the end and the beginning of the force protocol. This average energy difference is given by  $\Delta\langle E \rangle = \text{Tr}[H^H(\lambda(\tau)) - H(\lambda(0))] \rho_0$  and is also known as the untouched work (Talkner and Hänggi, 2016a). The difference between  $\langle w \rangle_\Lambda$  and  $\Delta\langle E \rangle$  vanishes only if the initial density matrix is diagonal with respect to the energy basis of the initial Hamiltonian. In general,  $\Delta\langle E \rangle - \langle w \rangle_\Lambda$  may be either positive or negative, and hence energy may seemingly be gained or lost in the TPPEMS if compared to the change of the average energies.<sup>24</sup> Attempts to interpret the energy mismatch in the spirit of Landauer's principle (Landauer, 1961) as equivalent to the gain of information proposed by Deffner, Paz, and Zurek (2016) and Kammerlander and Anders (2016), however, do not explain why the first energy measurement is energetically relevant but not the second one. Moreover, to translate information that can be quantified as negative Shannon entropy one needs to make contact with a thermal bath, even though the system is isolated during the entire force protocol. In particular, when the mismatch has a finite value because of a nonthermal initial state there is no natural choice to assign a temperature value and the information gain has no obvious energy equivalent.

Note that the characteristic function of work differs in form from one that specifies the statistics of an observable  $\mathcal{O}$ . In the latter case it had to take the form  $G(u) = \text{Tr} e^{i u \mathcal{O}} \rho$ . Hence, one cannot characterize work by an observable (Talkner, Lutz, and Hänggi, 2007). Yet the characteristic function of work [Eq. (90)] satisfies the formal sufficient and necessary conditions of being the Fourier transform of a probability density. These are (i)  $G_\Lambda(0) = 1$ , (ii)  $|G_\Lambda(u)| \leq 1$ , and (iii)  $\int dudv f^*(u) G_\Lambda(u - v) f(v) \geq 0$  for all integrable complex valued functions  $f(u)$  (Lukacs, 1970).<sup>25</sup>

<sup>24</sup>As a simple example, one may consider a two-level atom whose initial density matrix has diagonal elements  $p$  and  $1 - p$  and non-diagonal element  $q$  and  $q^*$  (the asterisk indicates complex conjugation) with  $p(1 - p) \geq |q|^2$  when specified in the energy eigenbasis of the initial Hamiltonian  $H_0$ . The force protocol consists of a sudden quench of the Hamiltonian with diagonal elements  $h_1$  and  $h_2$  and nondiagonal elements  $c$  and  $c^*$ , again with respect to the eigenbasis of  $H_0$ . The energy mismatch then becomes  $\Delta\langle E \rangle - \langle w \rangle = c q^* + c^* q$ , an expression that can take either sign.

<sup>25</sup>Condition (i) follows immediately with  $\text{Tr} \rho_0 = 1$ , condition (ii) is a consequence of  $|\text{Tr} A \rho| \leq \|A\|_B \|\rho_0\|_{TC}$ , where  $\|A\|_B = 1$  is the operator norm of the unitary operator  $A = e^{i u H^H(\lambda(\tau))} e^{-i u H(\lambda(0))}$  and  $\|\rho_0\|_{TC} = 1$  is the trace-class norm of the initial density matrix; for the definitions of the different operator norms see Schatten (1950). Finally, with  $C = \int dudv f(u) e^{-i u H^H(\lambda(\tau))} e^{i u H(\lambda(0))}$  condition (iii) follows according to  $\int dudv f^*(u) G_\Lambda(u - v) f(v) = \text{Tr} C^\dagger C \bar{\rho}_0 \geq 0$ .

Perarnau-Llobet *et al.* (2017) demonstrated that no measurement scheme<sup>26</sup> of work exists that is linear in the initial state of the system and for which the following two conditions are simultaneously satisfied: (1) The average work agrees with the difference of the average final and initial energies for any initial state, and (2) the resulting work statistics agrees for diagonal initial states (i.e.,  $\bar{\rho}_0 = \rho_0$ ) with the TPPEMS result [Eq. (89)]. A series of alternative attempts to define work in quantum systems other than by the TPPEMS were analyzed in view of this no-go theorem by Bäumer *et al.* (2018). We note that there exist two point measurement schemes using generalized energy measurements (Watanabe, Venkatesh, and Talkner, 2014) as well as generalized work measurements (Talkner and Hänggi, 2016a) for which it is possible to reconstruct the work distribution of the TPPEMS.

For systems initially in a Gibbs state  $\rho_0 = \rho(\beta, \lambda(0)) = Z^{-1}(\beta, \lambda(0))e^{-\beta H(\lambda(0))}$ , the TPPEMS leads to the quantum Jarzynski equality [Eq. (83)]. Likewise, the average of the irreversible work can be written as the Kullback-Leibler divergence between the actual final state and the Gibbs state at the initial temperature and final parameter values (Deffner and Lutz, 2010). Further, the average irreversible work  $\langle w_{\text{irr}} \rangle_\Lambda$  can be subdivided into a part that is due to coherences with respect to the final energy eigenbasis and another part that is caused by deviations of the finally reached populations of the final energy states from those of a canonical distribution with the final Hamiltonian at the initial temperature (Francica, Goold, and Plastina, 2019). Moreover, for Hamiltonians transforming under time reversal as  $\tilde{H}(\lambda) \equiv \theta H(\lambda) \theta^\dagger = H(\epsilon_\lambda \lambda)$ , the Crooks relation [Eq. (85)] is obeyed in exactly the same way as in classical systems (Tasaki, 2000; Talkner and Hänggi, 2007). Here  $\theta$  denotes the antiunitary time-reversal operator (Messiah, 1962) and  $\epsilon_\lambda$  is the parity of the parameter  $\lambda$  under time reversal. In general, for initial states differing from Gibbs states no fluctuation relations exist. Exceptions are microcanonical initial states for which a Crooks-type relation holds yet a Jarzynski equality is not known (Talkner, Hänggi, and Morillo, 2008; Talkner *et al.*, 2013). For grand canonical initial states both types of fluctuation relations hold. These relations involve both work and exchanged particle numbers together with the difference of the respective grand potentials (Yi, Kim, and Talkner, 2012).

We note that it is not possible to mutate the classical expression (81) into a quantum-mechanical form that is compatible with the fluctuation relations of Crooks and Jarzynski. A projective measurement of the *work operator*  $W = \int_0^\tau dt \mathcal{P}(t)$  defined in terms of a *power operator*  $\mathcal{P}(t) = \dot{\lambda}(t) \partial H^H(\lambda(t)) / \partial \lambda(t)$  yields on average the difference of the energy averages at the final and the initial times and, therefore, according to the findings of Perarnau-Llobet *et al.* (2017), cannot yield the *work statistics* of the TPPEMS for an initial state that is diagonal in the energy basis. Even the weaker requirement of satisfying the Jarzynski equality is not fulfilled

<sup>26</sup>We refer to a measurement scheme as a family of completely positive maps specifying the states after a selective measurement together with the probabilities of finding all possible results; for more details see, e.g., Chap. 2.4 of Breuer and Petruccione (2002).

(Engel and Nolte, 2007). In addition, a continuous weak measurement of the power operator  $\mathcal{P}(t)$  turns out to be incompatible with the fluctuation theorems (Venkatesh, Watanabe, and Talkner, 2015).

Finally, we remark that for a two point generalized energy measurement scheme the requirement that the Crooks relation is satisfied already restricts the allowed types of measurements to projective ones for systems with an infinite-dimensional Hilbert space. For systems with finite-dimensional Hilbert spaces slightly more general measurements are possible; the measurements still need to be error free, meaning that if the state in which the system is measured is an eigenstate belonging to a particular energy value, this energy value must be detected with certainty. For further details and the restrictions imposed by the Jarzynski equality we refer to the literature (Watanabe *et al.*, 2014; Ito *et al.*, 2019).

## B. Work in open systems

The work applied to an open system, which is part of a large closed system described by a Hamiltonian, as specified in Eq. (17), agrees with the work done on the total system only if system parameters are changed that influence neither the interaction nor the bath Hamiltonian. If the latter condition is not fulfilled, the work done on the total system can be identified with the change of the total system but the work done on the open system cannot be defined.<sup>27</sup>

We start with a discussion of work in open quantum systems and later specialize to the respective classical case.

### 1. Work in open quantum systems

The statistics of work performed on an open system upon changing a system parameter  $\lambda(t)$  according to a specified protocol  $\Lambda$  is formally determined by the same expression [Eq. (89)] as in the case of a closed system where all quantities refer to the total system (Campisi, Talkner, and Hänggi, 2009b). Specifically, in Eq. (89),  $E_m(t)$  indicates the eigenvalue of the total system Hamiltonian  $H_{\text{tot}}(\lambda(t)) = H_S(\lambda(t)) + H_B + H_{SB}$ . Likewise, the time-evolution operator

$$U_\Lambda = \mathcal{T} e^{i \int_0^\tau dt H_{\text{tot}}(\lambda(t)) / \hbar} \quad (91)$$

in Eq. (90) is governed by the total Hamiltonian, and the density matrix  $\rho_0$  specifies the total initial state. For a canonical initial state of the total system  $e^{-\beta H_{\text{tot}}(\lambda(0))} / \text{Tr} e^{-\beta H_{\text{tot}}(\lambda(0))}$  the fluctuation relations of Crooks and Jarzynski follow, with the free energy difference  $\Delta F = \Delta F_S$  holding because of  $F_S(\beta, \lambda) = F_{\text{tot}}(\beta, \lambda) - F_B(\beta)$  [see Eq. (34)] and the fact that the bare bath free energy is independent of the system parameter  $\lambda$ .<sup>28</sup> Hence, one has

<sup>27</sup>If, for example, both the system and the environment are electrically polarizable, the change of an externally controlled electromagnetic field directly affects both constituents.

<sup>28</sup>We note that the backward protocol also requires the reversal of all parameters transforming oddly under time reversal like magnetic fields, even if they are kept constant during the protocol or affect only the bath dynamics.

$$p_{\Lambda}(w) = e^{-\beta(\Delta F_S - w)} p_{\bar{\Lambda}}(-w), \quad (92)$$

$$\langle e^{-\beta w} \rangle_{\Lambda} = e^{-\beta \Delta F_S}, \quad (93)$$

i.e., the fluctuation relations continue to hold for open quantum systems that start in a total canonical equilibrium state independently of the coupling strength between system and bath and also irrespective of the nature of the open system's dynamics, with work and free energy differences both relating to the open system (Campisi, Talkner, and Hänggi, 2009b). This fact has raised doubts as to whether a work statistics within the TPPEMS contains any quantum aspects at all. These doubts have been removed by several case studies (Deffner and Lutz, 2008; Talkner, Burada, and Hänggi, 2008; Yi, Talkner, and Campisi, 2011; Yi, Kim, and Talkner, 2012) and also by the identification of quantum coherences generated during a force protocol (Blattmann and Mølmer, 2017; Miller and Anders, 2018; Francica, Goold, and Plastina, 2019) and by the investigation of quantum mechanically generated deviations of the work statistics from their classical Gaussian form for almost quasistatic isothermal processes (Bäumer *et al.*, 2019).

Yet in an experiment projective measurements of the total system Hamiltonian are to be imposed. Not only that they are difficult to perform, but the generally small difference between the much larger energies of the final and the initial state of the total system must be considered a severe practical limitation of the TPPEMS.

## 2. Work in open classical systems

The arguments leading to the quantum fluctuation relations for open systems can be repeated almost literally for classical open systems (Jarzynski, 2004). The dynamics of the classical total system is governed by a Hamiltonian of the form

$$H_{\text{tot}}(\mathbf{z}, \lambda(t)) = H_S(\mathbf{x}, \lambda) + H_B(\mathbf{y}) + H_{SB}(\mathbf{z}), \quad (94)$$

where  $\mathbf{z} = (\mathbf{x}, \mathbf{y})$  indicates a point in the phase space of the total system with components  $\mathbf{x}$  and  $\mathbf{y}$  specifying phase-space points of the system and its environment, respectively. Again, only the system Hamiltonian  $H(\mathbf{x}, \lambda)$  depends on the parameters  $\lambda$  that are subject to the protocol  $\Lambda$ . In analogy to Eqs. (79) and (81) the work can be expressed either as the energy difference of the total system or as an integral of the power to yield

$$w = H_{\text{tot}}(\mathbf{Z}(\mathbf{z}, \tau), \lambda(\tau)) - H_{\text{tot}}(\mathbf{z}, \lambda(0)) \quad (95)$$

$$= \int_0^{\tau} dt \frac{\partial H_S(\mathbf{X}(\mathbf{z}, t), \lambda(t))}{\lambda(t)} \dot{\lambda}(t), \quad (96)$$

where  $\mathbf{Z}(\mathbf{z}, t)$  denotes the trajectory in the full phase space starting at  $\mathbf{z}$  and  $\mathbf{X}(\mathbf{z}, t)$  is the projection of  $\mathbf{Z}(\mathbf{z}, t)$  onto the phase space of the open system. Therefore, for classical open systems measuring the total energy can be circumvented. Instead, the system trajectories during the protocol have to be monitored and used to calculate the supplied power. Hence, as for closed systems, based on the power supplied during the

protocol, the work done on an open system can be determined from the sole observation of the system trajectories.

In correspondence with Eq. (82), the work pdf becomes

$$p_{\Lambda}(w) = \int d\mathbf{z} \delta\left(w - \int_0^{\tau} dt \frac{\partial H_S(\mathbf{X}(\mathbf{z}, t), \lambda(t))}{\partial \lambda(t)} \dot{\lambda}(t)\right) \times \rho_0(\mathbf{z}), \quad (97)$$

with the total phase-space pdf  $\rho_0(\mathbf{z})$  characterizing the initial state. For total systems initially staying in a canonical state, the Jarzynski and Crooks relations follow (Jarzynski, 2004). As in the quantum case, the fluctuation relations hold for open classical systems irrespective of the kind of stochastic dynamics of the open system. For Markovian processes the fluctuation relations are derived directly from Fokker-Planck equations (Kurchan, 1998; Hatano and Sasa, 2001; Hummer and Szabo, 2001) and master equations (Jarzynski, 1997b; Gaspard, 2004; Harris and Schütz, 2007; Esposito, Harbola, and Mukamel, 2009).

## V. FLUCTUATING THERMODYNAMIC POTENTIALS

In this section we first restrict ourselves to a discussion of fluctuating thermodynamic potentials in thermal equilibrium for classical systems and only later comment on quantum mechanics. Inspired by the fact that the work performed on a system is a fluctuating quantity, one may ask whether it would not also be possible and even meaningful to consider fluctuating heat. Assuming the validity of an instantaneous first law one may construct from the fluctuating work and heat a likewise fluctuating internal energy as proposed in stochastic energetics (Sekimoto, 1998, 2010). Additional fluctuating thermodynamic potentials, in particular, fluctuating entropy, are considered in stochastic thermodynamics (Seifert, 2005, 2012; Van den Broeck and Esposito, 2015).

### A. Fluctuating internal energy

For a classical system in contact with its environment a fluctuating internal energy, as it is postulated using stochastic energetics (Sekimoto, 2010) and stochastic thermodynamics (Seifert, 2012), is supposed to assign to each momentary state  $\mathbf{x}$  of the open system a uniquely defined energy value. In general, one might expect such an assignment to also require some information about the actual state of the environment.<sup>29</sup> In the sequel we therefore consider the more general hypothesis that a fluctuating internal energy can be characterized by a function  $e(\mathbf{z}, \beta, \lambda)$ , where, as introduced in Sec. IV.B.2, points

<sup>29</sup>The dependence of the energy of an open system on the instantaneous state of the environment can be illustrated by the example of a dipolar molecule in a polar fluid. The magnitude and orientation of the molecule's electrical dipole moment relative to the local electric field determine a contribution to the energy of the molecule. Because the state of the fluid is not static in thermal equilibrium, and hence the orientation and magnitude of the local electrical field produced by the fluid surrounding the molecule fluctuate, an environmentally state-dependent, and therefore random, contribution to the molecule's energy results.

in the phase space of the open system are denoted by  $\mathbf{z} = (\mathbf{y}, \mathbf{x})$  with an environmental component  $\mathbf{y}$  and an open system component  $\mathbf{x}$ . Because the microstate  $\mathbf{y}$  of the environment is not monitored, the hypothetical fluctuating internal energy  $e(\mathbf{z}, \beta, \lambda)$  has to be considered a random field, where the random variable  $\mathbf{y}$  is distributed according to the conditional pdf  $w(\mathbf{y}|\mathbf{x})$  defined in Eq. (32). A basic requirement for a fluctuating internal energy is that its average with respect to the canonical equilibrium state of the total system must coincide with the internal energy  $U_S$  of the total system so that with  $\rho_{\text{tot}}(\mathbf{z}, \beta, \lambda)$  in Eq. (18)

$$U_S(\beta, \lambda) = \int d\mathbf{z} e(\mathbf{z}, \beta, \lambda) \rho_{\text{tot}}(\mathbf{z}, \beta, \lambda). \quad (98)$$

Combined with Eq. (37) for the internal energy one finds that any fluctuating internal energy must be of the form

$$\begin{aligned} e(\mathbf{z}, \beta, \lambda) &= \frac{\partial}{\partial \beta} [\beta H^*(\mathbf{x}, \beta, \lambda)] + h_u(\mathbf{z}, \beta, \lambda) \\ &= \langle H_{\text{tot}} | \mathbf{x} \rangle - \langle H_B \rangle_B + h_u(\mathbf{z}, \beta, \lambda), \end{aligned} \quad (99)$$

where  $h_u(\mathbf{z}, \beta, \lambda) \in \mathcal{N}_{\beta, \lambda}$  is a random field with vanishing mean value in thermal equilibrium. Accordingly, the set  $\mathcal{N}_{\beta, \lambda} = \{h(\mathbf{z}) | \int d\mathbf{z} h(\mathbf{z}) \rho_{\text{tot}}(\mathbf{z}, \beta, \lambda) = 0\}$  consists of all random fields with vanishing equilibrium average for fixed  $\rho_{\text{tot}}(\mathbf{z}, \beta, \lambda)$ . The second line, in which the fluctuating internal energy is expressed as the surplus of the conditional total energy relative to the bare environmental energy superimposed by a fluctuating contribution  $h_u(\mathbf{z}, \beta, \lambda)$ , is obtained with the help of Eq. (31).<sup>30</sup> It is worth noting that with the subtraction of the average bare environment energy one may assign a finite energy to the open system even for large environments in the thermodynamic limit. Without this term, the fluctuating internal energy would depend in a sensitive way on irrelevant details of the environment. Moreover, without it, the average fluctuating energy of the open system would be given by the internal energy of the total system rather than by the internal energy difference of the total system and the bare environment as required by Eq. (34).<sup>31</sup>

Once a fluctuating internal energy is assigned to the state of the total system, the momentary energy content  $g(\mathbf{z}, \beta, \lambda)$  of

<sup>30</sup>One possible choice of a random field satisfying  $h_u(\mathbf{z}, \beta, \lambda) \in \mathcal{N}_{\beta, \lambda}$  is given by  $h_u(\mathbf{z}, \beta, \lambda) = \alpha(\mathbf{x}) \delta H_{SB}(\mathbf{z})$ , assigning an arbitrary fraction  $\alpha(\mathbf{x})$  of the fluctuation of the interaction energy  $\delta H_{SB}(\mathbf{z}) = H_{SB}(\mathbf{z}) - \langle H_{SB}(\mathbf{z}) | \mathbf{x} \rangle$  as an internal energy fluctuation (Talkner and Hänggi, 2016b).

<sup>31</sup>At variance with the previous definition of the fluctuating internal energy by Seifert (2016) that corresponds to the choice with a vanishing random field  $h_u(\mathbf{z}, \beta, \lambda) = 0$  in Eq. (99), in a recent review of stochastic thermodynamics (Seifert, 2019) the fluctuating internal energy is proposed to agree with the conditional total system energy  $\langle H_{\text{tot}} | \mathbf{x} \rangle$ , i.e., without the subtraction of the bare bath energy. Hence, small systems also acquire the typically large, possibly even diverging energy of the environment.

the reservoir can be identified as the difference of the total energy and the fluctuating internal energy, yielding

$$g(\mathbf{z}, \beta, \lambda) = H_{\text{tot}}(\mathbf{z}, \lambda) - e(\mathbf{z}, \beta, \lambda). \quad (100)$$

With this assignment one may define the heat  $q$  exchanged with the environment in the course of a process in which a system-parameter change from  $\lambda$  to  $\lambda'$  leads the total system to move in phase space from the initial point  $\mathbf{z}$  to the final point  $\mathbf{z}'$  as

$$q = g(\mathbf{z}, \beta, \lambda) - g(\mathbf{z}', \beta, \lambda'). \quad (101)$$

With this definition, a positive heat corresponds to an energy taken from the environment. The work  $w$ , which is performed on the open system in the same realization of the process, is, according to Eq. (95), given by the difference of the total Hamiltonians and hence written as

$$w = H_{\text{tot}}(\mathbf{z}', \lambda') - H_{\text{tot}}(\mathbf{z}, \lambda). \quad (102)$$

The fluctuating work, heat, and internal energy then satisfy the first-law-like relation

$$\Delta e = q + w, \quad (103)$$

which, however, is of little predictive power because both the fluctuating internal energy change and the heat depend on the difference of  $h_u(\mathbf{z}, \beta, \lambda)$  and  $h_u(\mathbf{z}', \beta, \lambda')$ . These are values of almost arbitrary functions, which are restricted only by having vanishing equilibrium averages. Therefore, only if both the initial and final states of a system under the influence of forcing are equilibrium states, according to Eq. (99), the dependence of the average internal energies on the functions  $h_u(\mathbf{z}, \beta, \lambda(t))$ ,  $t = 0, \tau$  disappears and consequently the average of heat can be determined from the difference of the final and initial internal energies and the average work done by the force. More can be done only in the weak coupling limit; see Sec. V.D.

Even though, as mentioned at the start of Sec. III, the properties of the open system, including its dynamics, are identical for a canonical ensemble of large closed systems and, for a single large system that weakly couples to a superbath at the required temperature, the present definition of heat is restricted to the former situation because otherwise after a sufficiently large time the heat produced in a cyclic process is finally absorbed by the superbath; see Talkner and Hänggi (2016b), particularly footnote 8 therein.

## B. Fluctuating entropy and free energy

Once a particular fluctuating internal energy is chosen, fluctuating free energies  $f(\mathbf{z}, \beta, \lambda)$  and entropies  $s(\mathbf{z}, \beta, \lambda)$  may be assigned under the constraint that their equilibrium averages coincide with the respective potentials  $F_S(\beta, \lambda)$  and  $S_S(\beta, \lambda)$  of the open system such that

$$F_S(\beta, \lambda) = \int d\mathbf{z} f(\mathbf{z}, \beta, \lambda) \rho_{\text{tot}}(\mathbf{z}, \beta, \lambda), \quad (104)$$

$$S_S(\beta, \lambda) = \int d\mathbf{z} s(\mathbf{z}, \beta, \lambda) \rho_{\text{tot}}(\mathbf{z}, \beta, \lambda), \quad (105)$$

where the pdf  $\rho_{\text{tot}}(\mathbf{z}, \beta, \lambda)$  is given by Eq. (18). To obtain a thermodynamically consistent description of the open system we require the validity of Eqs. (12) and (13) between the open system's thermodynamic potentials  $U_S$  and  $S_S$ , respectively, and the corresponding free energy  $F_S$ , yielding

$$\int d\mathbf{z} \rho_{\text{tot}}(\mathbf{z}, \beta, \lambda) \left\{ e(\mathbf{z}, \beta, \lambda) - \frac{\partial}{\partial \beta} \beta f(\mathbf{z}, \beta, \lambda) - \beta f(\mathbf{z}, \beta, \lambda) \frac{\partial}{\partial \beta} \ln \rho_{\text{tot}}(\mathbf{z}, \beta, \lambda) \right\} = 0, \quad (106)$$

$$\int d\mathbf{z} \rho_{\text{tot}}(\mathbf{z}, \beta, \lambda) \left[ s(\mathbf{z}, \beta, \lambda) - k_B \beta^2 \left( \frac{\partial}{\partial \beta} f(\mathbf{z}, \beta, \lambda) + f(\mathbf{z}, \beta, \lambda) \frac{\partial}{\partial \beta} \ln \rho_{\text{tot}}(\mathbf{z}, \beta, \lambda) \right) \right] = 0. \quad (107)$$

Accordingly, one obtains for the fluctuating thermodynamic potentials the consistency relations

$$e(\mathbf{z}, \beta, \lambda) = \frac{\partial}{\partial \beta} \beta f(\mathbf{z}, \beta, \lambda) - \beta f(\mathbf{z}, \beta, \lambda) (H_{\text{tot}}(\mathbf{z}, \lambda) - U_{\text{tot}}(\beta, \lambda)) + h_e(\mathbf{z}, \beta, \lambda), \quad (108)$$

$$s(\mathbf{z}, \beta, \lambda) = k_B \beta^2 \left( \frac{\partial}{\partial \beta} f(\mathbf{z}, \beta, \lambda) - f(\mathbf{z}, \beta, \lambda) (H_{\text{tot}}(\mathbf{z}, \lambda) - U_{\text{tot}}(\beta, \lambda)) \right) + h_s(\mathbf{z}, \beta, \lambda), \quad (109)$$

where the right-hand sides of Eqs. (108) and (109) contain arbitrary functions  $h_{e/s}(\mathbf{z}, \beta, \lambda) \in \mathcal{N}_{\beta, \lambda}$ . Apparently, the requirement of thermodynamic consistency is not sufficient to assess these functions other than by mere definitions.

Making the assumption that the fluctuating thermodynamic potentials do not explicitly depend on the environmental variables  $\mathbf{y}$  one is left with consistency conditions of the same type as Eqs. (108) and (109) in which the full phase-space variable  $\mathbf{z}$  is replaced by  $\mathbf{x}$ . Moreover, in both equations the expression in parentheses has to be modified according to  $(H_{\text{tot}}(\mathbf{z}, \lambda) - U_{\text{tot}}(\beta, \lambda)) \rightarrow \partial(\beta[H^*(\mathbf{x}, \beta, \lambda) - F_S(\beta, \lambda)])/\partial\beta$  (Talkner and Hänggi, 2016b). The unknown functions  $h_{u/e/s}(\mathbf{z}, \beta, \lambda) \in \mathcal{N}_{\beta, \lambda}$  of the total phase space must then be replaced by functions  $h_{u/e/s}(\mathbf{x}, \beta, \lambda)$  that have vanishing average with respect to the reduced equilibrium pdf  $\rho_S(\mathbf{x}, \beta, \lambda) = \int d\mathbf{y} \rho_{\text{tot}}(\mathbf{z}, \beta, \lambda)$ .

It might be tempting to choose the functions  $h_e(\mathbf{x}, \beta, \lambda)$  and  $h_s(\mathbf{x}, \beta, \lambda)$  in such a way that the fluctuating potentials satisfy the same relations as the respective average quantities do and therefore should be related by (i)  $e(\mathbf{x}, \beta, \lambda) = \partial(\beta f(\mathbf{x}, \beta, \lambda))/\partial\beta$  and (ii)  $s(\mathbf{x}, \beta, \lambda) = k_B \beta^2 \partial(f(\mathbf{x}, \beta, \lambda))/\partial\beta$  (Seifert, 2019). To see whether this assumption is compatible with the required relations for the averages, both sides are

averaged with respect to the open system equilibrium pdf  $\rho_S(\mathbf{x}, \beta, \lambda)$ . From the first equation one obtains  $U_S(\beta, \lambda) = \int d\mathbf{x} \rho_S(\mathbf{x}, \beta, \lambda) \partial(\beta f(\mathbf{x}, \beta, \lambda))/\partial\beta$ . Therefrom, together with the thermodynamic consistency equation (12) the condition  $\int d\mathbf{x} f(\mathbf{x}, \beta, \lambda) \partial\rho(\mathbf{x}, \beta, \lambda)/\partial\beta = 0$  follows, which, in general, does not hold. The second equation yields the same condition on the fluctuating free energy. This implies that the thermodynamic consistency is violated, in general, by both relations (i) and (ii).<sup>32</sup> Strasberg and Esposito (2017) obtained the same inconsistent relations (i) and (ii) using an approximate coarse graining procedure of a master equation.

### C. Fluctuating work and heat in open quantum systems

According to the detailed discussion in Sec. IV.B, work can be understood as the difference of the results of two energy measurements of the total system. To obtain an analogous definition of quantum heat, one needs to know a convenient operator  $g$  representing the energy content of the environment. Then the difference of the outcomes of two projective measurements of this operator yields the heat, i.e., the energy lost by the environment.

For processes during which the system is alternately coupled to and decoupled from environments,<sup>33</sup> the environmental energy is determined by the Hamiltonian  $H_B$  of the bare environment. In other situations with a permanent contact of system and environment one may follow the strategy for classical systems based on a fluctuating internal energy, as outlined in Sec. V.A. The quantum analog of a fluctuating internal energy is an internal energy operator  $e(\beta, \lambda)$  with the property to yield the internal energy  $U_S$  on average in thermal equilibrium, i.e.,  $U_S(\beta, \lambda) = \text{Tr}_{\text{tot}}[e(\beta, \lambda) \rho_{\text{tot}}(\beta, \lambda)] = \text{Tr}_{\text{tot}}[\partial\beta H^*(\beta, \lambda)/\partial\beta \rho_{\text{tot}}(\beta, \lambda)]$ . As in the classical case, this requirement leaves considerable ambiguity as to the choice of an internal energy operator, which can be represented as

$$e(\beta, \lambda) = \frac{\partial}{\partial \beta} \beta H^*(\beta, \lambda) + h_e(\beta, \lambda), \quad (110)$$

where  $h_e(\beta, \lambda)$  is a Hermitian operator with vanishing equilibrium average  $\text{Tr}_{\text{tot}} h_e(\beta, \lambda) \rho_{\text{tot}}(\beta, \lambda) = 0$ . To any internal energy operator there belongs a corresponding operator  $g(\beta, \lambda)$  specifying the energy content of the environment with

$$g(\beta, \lambda) = H_{\text{tot}} - e(\beta, \lambda). \quad (111)$$

The heat characterizing a particular process can then be operationally defined in terms of two measurements of this

<sup>32</sup>The thermodynamically consistent fluctuating internal energy and entropy postulated by Seifert (2016) imply a deterministic free energy  $f(\mathbf{z}, \beta, \lambda) = F_S(\beta, \lambda)$  (Talkner and Hänggi, 2016b).

<sup>33</sup>Such situations are realized in heat and particle exchange between reservoirs (Andrieux *et al.*, 2009; Campisi, Talkner, and Hänggi, 2010; Campisi, Hänggi, and Talkner, 2011; Jeon *et al.*, 2017) and also in cyclically performing engines (Zheng, Hänggi, and Poletti, 2016; Kosloff and Rezek, 2017; Ding *et al.*, 2018). Possible changes of the energy of the total system due to the time dependence of the coupling and decoupling are typically neglected.

environmental energy operator at the beginning and the end of the respective processes. As in the classical case the environmental energy  $g(\beta, \lambda)$  and, consequently, the heat inherit the ambiguity of the internal energy operator.

If one is interested in the amount of work and heat that is concurrently supplied to the system and exchanged with the environment in the same process, one faces the problem of having to simultaneously measure the Hamiltonian of the total system and environmental energy operator  $g(\beta, \lambda)$ . These operators do not, in general, commute with each other in the presence of an interaction between system and environment. Hence, for systems continuously in contact with their environment a process cannot be characterized by a simultaneous specification of work and heat for the same reason that position and momentum cannot be assigned to a quantum particle at the same time. The only exception to this rule is realized for a system weakly coupling with its environment, as discussed in more detail later. The formulation of a first law for systems other than weakly interacting quantum ones therefore seems doubtful to us, contrary to a widespread opposing opinion (Alicki, 1979; Nieuwenhuizen and Allahverdyan, 2002; Seifert, 2016).

#### D. Weak coupling

As mentioned in Sec. III.A, in the weak coupling limit (Van Hove, 1957; Davies, 1976) one considers an interaction of vanishingly small strength  $\kappa$  between system and environment acting on an increasingly long timescale such that energy may still flow between system and environment and eventually the small system equilibrates without a noticeable renormalization of the system's Hamiltonian. Technically speaking, the internal energy operator of the system then agrees with the bare system Hamiltonian  $e(\beta, \lambda) = H_S(\lambda)$  and the environmental energy operator with its bare Hamiltonian  $g(\beta, \lambda) = H_B$  (Talkner, Campisi, and Hänggi, 2009). Additional small contributions resulting from the interaction can be neglected in any respect other than for the long time dynamics. This results in the exceptional situation in which quantum work and quantum heat can be determined for the same process by simultaneously measuring  $H_{\text{tot}}(\lambda) = H_S(\lambda) + H_B$  and  $g(\beta, \lambda)$ , yielding

$$w = E_{m'}(\tau) + \epsilon_{\alpha'} - [E_m(0) + \epsilon_{\alpha}], \quad (112)$$

$$q = \epsilon_{\alpha} - \epsilon_{\alpha'}, \quad (113)$$

where  $E_m(t)$  denotes an eigenvalue of the system Hamiltonian  $H_S(\lambda(t))$ ,  $\epsilon_{\alpha}$  denotes the eigenvalue of  $H_B$  emerging in the first measurement, and  $\epsilon_{\alpha'}$  denotes the corresponding result of the second measurement. The sum of heat and work is given by the difference of eigenvalues of the system Hamiltonian  $H_S$ , which is consistent with  $e(\beta, \lambda) = H_S(\lambda)$ .

For a force protocol  $\Lambda$  extending over the time span  $(0, \tau)$  the joint work and heat pdf  $p_{\Lambda}(w, q)$  becomes (Talkner, Campisi, and Hänggi, 2009)

$$p_{\Lambda}(w, q) = \sum_{\substack{m, m' \\ \alpha, \alpha'}} \delta(w - E_{m'} - \epsilon_{\alpha'} + E_m + \epsilon_{\alpha}) \times \delta(q - \epsilon_{\alpha} + \epsilon_{\alpha'}) p_{\Lambda}(m', \alpha'; m, \alpha). \quad (114)$$

Here  $p_{\Lambda}(m', \alpha'; m, \alpha)$  specifies the joint probability of finding the total system at the energy  $E_m(0)$  and the environment at  $e_{\alpha}$  immediately before the force protocol starts and at  $E_{m'}(\tau)$  and  $\epsilon_{\alpha'}$  at the end. It can be written as  $p_{\Lambda}(m', \alpha'; m, \alpha) = p_{\Lambda}(m', \alpha' | m, \alpha) p(m, \alpha)$  in terms of the initial probability distribution of the total system  $p(m, \alpha) = \text{Tr} \Pi_m(0) Q_{\alpha} \rho_{\text{tot}}$  and the transition probability  $p_{\Lambda}(m', \alpha' | m, \alpha) = \text{Tr}_{\text{tot}} \Pi_{m'}(\tau) Q_{\alpha'} U_{\Lambda} \Pi_m(0) Q_{\alpha} U_{\Lambda}^{\dagger} / \text{Tr}_{\text{tot}} \Pi_m(0)$ . The projection operators onto the eigenspaces of the Hamiltonians  $H_S(\lambda(t))$  and  $H_B$  are denoted by  $\Pi_n(t)$  and  $Q_{\alpha}$ , respectively. The time-evolution operator

$$U_{\Lambda} = \mathcal{T} e^{-i \int_0^{\tau} dt H_{\text{tot}} / \hbar} \quad (115)$$

is governed by the full Hamiltonian of the total system, including the interaction. For short processes of duration  $\tau$  with  $\kappa^2 \tau \ll 1$  the environmental dynamics is unaffected by the interaction and hence, with  $\epsilon_l = \epsilon_k$ , the heat typically vanishes. Bath transitions and, accordingly, the heat transfer become important for long-lasting processes with  $\kappa^2 \tau \gtrsim 1$ .

The joint work and heat pdf, Eq. (114), describing a process controlled by the force protocol  $\Lambda$  and starting from a canonical equilibrium state of the total system at the inverse temperature  $\beta$  is linked to the according pdf for the reversed protocol  $\bar{\Lambda}$  by a Crooks-type relation (Talkner, Campisi, and Hänggi, 2009) given as

$$p_{\Lambda}(w, q) = e^{-\beta(\Delta F_S - w)} p_{\bar{\Lambda}}(-w, -q). \quad (116)$$

As an immediate consequence one recovers for the marginal work pdf  $p_{\Lambda}(w) = \int dq p_{\Lambda}(w, q)$  the Crooks relation [Eq. (92)] and the Jarzynski equality [Eq. (93)] for open systems. In contrast, the marginal heat pdf  $p_{\Lambda}^q(q) = \int dw p_{\Lambda}(w, q)$  does not obey a fluctuation relation. We further note that the two fluctuation theorems hold only for the work defined as the energy difference of the total system. In contrast, the joint pdf  $p_{\Lambda}^{\Delta e, q}(\Delta e, q) = \int dw \delta(\Delta e - w - q) p_{\Lambda}(w, q)$  of the difference of the according internal energy  $\Delta e = E_{m'}(\tau) - E_m(0)$  and of the heat satisfies a Crooks-type relation of the form (Talkner, Campisi, and Hänggi, 2009)

$$p_{\Lambda}^{\Delta e, q}(\Delta e, q) = e^{-\beta(\Delta F - \Delta e + q)} p_{\bar{\Lambda}}^{\Delta e, q}(-\Delta e, -q). \quad (117)$$

But because of the presence of the heat in the exponent on the right-hand side, one does not obtain a Jarzynski equality in  $\Delta e$ , other than for sufficiently short protocols for which the heat vanishes but decoherence may already take place (Smith *et al.*, 2018).

Independent of how strong the interaction between a system and its environment is, the time rate of change of the average bare energy  $E_S(t) = \text{Tr}_{SB} H_S(t) \rho_{\text{tot}} = \text{Tr}_S H_S(t) \rho_S(t)$  can always be split into two contributions according to

$$\dot{E}_S(t) = \text{Tr}_S \frac{\partial H_S(t)}{\partial t} \rho_S(t) + \text{Tr}_S H_S(t) \dot{\rho}_S(t). \quad (118)$$

Equation (118) can be considered only as a proper formulation of the first law in the weak coupling limit. Only then does the internal energy of the open system coincide with the thermal average of the bare system Hamiltonian. Independent of the coupling strength the integral of the first term on the right-hand side, extended over the time span of the force protocol, yields the average work done on the system according to the TPPEMS provided that the initial state  $\rho_{\text{tot}0}$  and the initial total Hamiltonian  $H_{\text{tot}}(\lambda(0))$  commute with each other. In general, the integrated second term, however, has no definite physical meaning.

Beyond the weak coupling limit the lack of a uniquely defined fluctuating internal energy makes it impossible even to assign an average internal energy to anything other than the according thermal equilibrium state. Because, in general, at the end of a force protocol the system is not in an equilibrium state, it is therefore not possible to assign the change of the average internal energy in the respective process. Consequently, an average heat also cannot be specified.

## VI. OUTLOOK AND CONCLUSIONS

The central notion in the thermodynamics of open systems staying in strong contact with the environment is given by the Hamiltonian of mean force, which is defined in terms of an average of the Boltzmann factor over the thermally distributed environmental degrees of freedom. It provides at the same time the reduced density matrix of the open system and its thermodynamic equilibrium properties, which, in general, are influenced by the environment in a way that they cannot be inferred from the sole knowledge of the reduced density matrix. Owing to the fact that the resulting partition function  $Z_S$  of the open system is given by the ratio of the partition functions of the total system and the bare environment, the existence of  $Z_S$  and its independence of irrelevant details of the environment is guaranteed. In particular, remote parts of the environment coupling only weakly to the system do not affect  $Z_S$ . A further important consequence of this particular structure is the finding that the thermodynamic potentials as well as all of their derivatives relating to the open system are determined by differences of the respective quantities referring to the total system and the environment. This guarantees the thermodynamic consistency of the thermodynamic potentials and the validity of the third law. It also may exhibit unusual properties like negative entropy and negative specific heat without, however, indicating any instabilities of the respective systems. In the case of negative entropy it indicates that the interaction between system and environment enforces a state with a higher order than in its absence.

The attempt to represent the thermodynamic internal energy of an open system as an equilibrium average of a fluctuating internal energy in the case of classical systems, or, for quantum systems, as an internal energy operator, leads to a tremendous ambiguity in the choice of these fluctuating or operator-valued internal energy expressions. Other fluctuating potentials like fluctuating entropy and fluctuating free energy, as well as the corresponding quantum-mechanical operator-valued expressions, are also

affected by these ambiguities. The interpretation of this inconclusiveness as a kind of gauge freedom (Jarzynski, 2017) seems rather far-fetched. Other than in proper gauge theories there is no obvious advantage to considering gauge-dependent quantities in the present context. The fact that a fluctuating thermodynamic potential, on the one hand, plays the role of an observable but, on the other hand, depends on the Gibbs state of the total system appears to be a strange mixture of the two fundamentally distinct categories of states and observables.

Because the specification of heat relies on the division of the internal energy in work and heat, the notion of heat inherits the ambiguity of the fluctuating internal energy. While the work as a fluctuating quantity can be expressed in an experimentally accessible way for classical systems (Collin *et al.*, 2005), in quantum systems the TPPEMS of the total energy presents a major experimental challenge (An *et al.*, 2015).<sup>34</sup> For quantum systems the concurrent determination of heat and work is additionally hampered by the fact that it relies on simultaneous measurements of two energy expressions that do not commute, except for systems weakly coupling to an environment. In the latter case, for systems that couple weakly to their environment the internal energy can be characterized by the bare system Hamiltonian, and the environmental energy is characterized by its bare bath Hamiltonian.

Before closing we note that in this Colloquium we do not consider further relations between thermodynamics and information theory (Vinjanampathy and Anders, 2016; Strasberg *et al.*, 2017) other than those between the thermodynamic entropy of an open system and several information-theoretic notions in Sec. III.C, nor did we discuss the related recent resource theory approach (Chitambar and Gour, 2019). In this context we stress that the frequently made identification of information entropy, typically given by a Shannon or von Neumann entropy, with the thermodynamic entropy must be considered with utmost care as it is by no means guaranteed to be correct (Norton, 2013; Hänggi and Talkner, 2015; Hänggi, Hilbert, and Dunkel, 2016; Alicki and Horodecki, 2019). A few further aspects of classical systems outside of equilibrium were considered by Talkner and Hänggi (2016b).

## ACKNOWLEDGMENTS

We thank Michele Campisi, Gert-Ludwig Ingold, Juzar Thingna, Prasanna Venkatesh, Gentaro Watanabe, and Juyeon Yi for the numerous discussions and collaborations on the thermodynamics of open systems and fluctuation relations. Valuable discussions with Massimiliano Esposito, Christopher Jarzynski, Abraham Nitzan, Marti Perarnau-Llobet, Udo Seifert, and Philipp Strasberg are gratefully acknowledged.

<sup>34</sup>The experimental determination of the average work in terms of the average power already provides a major challenge as it requires a sufficiently frequent repetition of experimental runs to a variable upper time (Clos *et al.*, 2016).

## REFERENCES

- Abramowitz, M., and I. A. Stegun, 1964, Eds., *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, DC).
- Alicki, R., 1979, “The quantum open system as a model of the heat engine,” *J. Phys. A* **12**, L103.
- Alicki, R., and M. Horodecki, 2019, “Information-thermodynamics link revisited,” *J. Phys. A* **52**, 204001.
- An, S., J.-N. Zhang, M. Um, D. Lv, Y. Lu, J. Zhang, Z.-Q. Yin, H. T. Quan, and K. Kim, 2015, “Experimental test of the quantum Jarzynski equality with a trapped ion system,” *Nat. Phys.* **11**, 193.
- Andrieux, D., P. Gaspard, T. Monnai, and S. Tasaki, 2009, “The fluctuation theorem for currents in open quantum systems,” *New J. Phys.* **11**, 043014.
- Aurell, E., 2018, “Unified picture of strong-coupling stochastic thermodynamics and time reversals,” *Phys. Rev. E* **97**, 042112.
- Badii, R., E. Brun, M. Finardi, L. Flepp, R. Holzner, J. Parisi, C. Reyl, and J. Simonet, 1994, “Progress in the analysis of experimental chaos through periodic orbits,” *Rev. Mod. Phys.* **66**, 1389–1415.
- Baltes, H. P., and E. R. Hilf, 1976, *Spectra of Finite Systems* (Bibliographisches Institut, Mannheim, Germany).
- Bäumer, E., M. Lostaglio, M. Perarnau-Llobet, and R. Sampaio, 2018, “Fluctuating work in coherent quantum systems: Proposals and limitations,” in *Thermodynamics in the Quantum Regime*, edited by F. Binder, L. A. Correa, C. Gogolin, J. Anders, and G. Adesso (Springer Nature, Cham, Switzerland), p. 275.
- Bäumer, E., M. Perarnau-Llobet, P. Kammerlander, H. Wilming, and R. Renner, 2019, “Imperfect thermalization allow for optimal thermodynamic processes,” *Quantum* **3**, 153.
- Blattmann, R., and K. Mølmer, 2017, “Macroscopic realism of quantum work fluctuations,” *Phys. Rev. A* **96**, 012115.
- Bochkov, G. N., and Y. E. Kuzovlev, 1981a, “Nonlinear fluctuation-dissipation relations and stochastic models in nonequilibrium thermodynamics I. Generalized fluctuation-dissipation theorem,” *Physica (Amsterdam)* **106A**, 443.
- Bochkov, G. N., and Y. E. Kuzovlev, 1981b, “Nonlinear fluctuation-dissipation relations and stochastic models in nonequilibrium thermodynamics II. Kinetic potential and variational principles for nonlinear irreversible processes,” *Physica (Amsterdam)* **106A**, 480.
- Bogolyubov, N. N., 1945, *Some Statistical Methods in Mathematical Physics* (National Academy of Sciences of Ukraine, Kiev, Ukraine), p. 115 (in Russian).
- Bopp, F., 1953, “Übergang von der mikrokanonischen zur kanonischen Gesamtheit mittels  $\delta$ -Funktion,” *Z. Naturforsch. A* **8**, 233–234.
- Braginsky, V. B., Y. I. Vorontsov, and K. S. Thorne, 1980, “Quantum nondemolition measurements,” *Science* **209**, 547.
- Breuer, H.-P., and F. Petruccione, 2002, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford).
- Caldeira, A. O., and A. J. Leggett, 1983, “Quantum tunneling in a dissipative medium,” *Ann. Phys. (N.Y.)* **149**, 374–456.
- Callen, H. B., 1985, *Thermodynamics and an Introduction to Thermostatistics*, 2nd ed. (John Wiley, New York).
- Campisi, M., 2007, “On the limiting cases of nonextensive thermostatics,” *Phys. Lett. A* **366**, 335–338.
- Campisi, M., 2015, “Construction of microcanonical entropy on thermodynamic pillars,” *Phys. Rev. E* **91**, 052147.
- Campisi, M., P. Hänggi, and P. Talkner, 2011, “Colloquium: Quantum fluctuation relations: Foundations and applications,” *Rev. Mod. Phys.* **83**, 771–791.
- Campisi, M., P. Talkner, and P. Hänggi, 2009a, “Finite bath fluctuation theorem,” *Phys. Rev. E* **80**, 031145.
- Campisi, M., P. Talkner, and P. Hänggi, 2009b, “Fluctuation Theorem for Arbitrary Open Quantum Systems,” *Phys. Rev. Lett.* **102**, 210401.
- Campisi, M., P. Talkner, and P. Hänggi, 2009c, “Thermodynamics and fluctuation theorems for a strongly coupled open quantum system: An exactly solvable case,” *J. Phys. A* **42**, 392002.
- Campisi, M., P. Talkner, and P. Hänggi, 2010, “Fluctuation Theorems for Continuously Monitored Quantum Fluxes,” *Phys. Rev. Lett.* **105**, 140601.
- Campisi, M., P. Talkner, and P. Hänggi, 2011, “Quantum Bochkov-Kuzovlev work fluctuation theorems,” *Phil. Trans. R. Soc. A* **369**, 291.
- Campisi, M., D. Zueco, and P. Talkner, 2010, “Thermodynamic anomalies in an open quantum system: Strong coupling effects in the isotropic XY model,” *Chem. Phys.* **375**, 187.
- Castelvecchi, D., 2017, “Thermodynamics: Clash of the physics laws,” *Nature (London)* **543**, 597.
- Chitambar, E., and G. Gour, 2019, “Quantum resource theories,” *Rev. Mod. Phys.* **91**, 025001.
- Clos, G., D. Porras, U. Warring, and T. Schaetz, 2016, “Time-Resolved Observation of Thermalization in an Isolated Quantum System,” *Phys. Rev. Lett.* **117**, 170401.
- Collin, D., F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante, 2005, “Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies,” *Nature (London)* **437**, 231.
- Crooks, G. E., 1999, “Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences,” *Phys. Rev. E* **60**, 2721.
- Davies, E. B., 1976, *Quantum Theory of Open Systems* (Academic Press, London).
- Deffner, S., and E. Lutz, 2008, “Nonequilibrium work distribution of a quantum harmonic oscillator,” *Phys. Rev. E* **77**, 021128.
- Deffner, S., and E. Lutz, 2010, “Generalized Clausius Inequality for Nonequilibrium Processes,” *Phys. Rev. Lett.* **105**, 170402.
- Deffner, S., J. P. Paz, and W. H. Zurek, 2016, “Quantum work and the thermodynamic cost of quantum measurements,” *Phys. Rev. E* **94**, 010103.
- Deng, J. W., J. D. Jaramillo, P. Hänggi, and J. B. Gong, 2017, “Deformed Jarzynski equality,” *Entropy* **19**, 419.
- Deng, J. W., A. M. Tan, P. Hänggi, and J. B. Gong, 2017, “Merits and qualms of work fluctuations in classical fluctuation theorems,” *Phys. Rev. E* **95**, 012106.
- Deutsch, J. M., 1991, “Quantum statistical mechanics in a closed system,” *Phys. Rev. A* **43**, 2046.
- Ding, X. H., J. Yi, Y. W. Kim, and P. Talkner, 2018, “Measurement driven single temperature engine,” *Phys. Rev. E* **98**, 042122.
- Dou, W., M. A. Ochoa, A. Nitzan, and J. E. Subotnik, 2018, “Thermodynamics of a subensemble of a canonical ensemble,” *Phys. Rev. B* **98**, 134306.
- Duarte, O. S., and A. O. Caldeira, 2006, “Effective Coupling between Two Brownian Particles,” *Phys. Rev. Lett.* **97**, 250601.
- Dunkel, J., and S. Hilbert, 2006, “Phase transitions in small systems: Microcanonical vs. canonical ensembles,” *Physica (Amsterdam)* **370A**, 390–406.
- Dunkel, J., and S. Hilbert, 2014, “Consistent thermostatics forbids negative absolute temperatures,” *Nat. Phys.* **10**, 67–72.
- Ellis, R. S., 1985, *Entropy, Large Deviations, and Statistical Mechanics*, Grundlehren der mathematischen Wissenschaften Vol. 271 (Springer-Verlag, Berlin).
- Engel, A., and R. Nolte, 2007, “Jarzynski equation for simple quantum systems: Comparing two definitions of work,” *Europhys. Lett.* **79**, 10003.

- Esposito, M., U. Harbola, and S. Mukamel, 2009, “Nonequilibrium fluctuations, fluctuation theorems, and counting statistics,” *Rev. Mod. Phys.* **81**, 1665.
- Ford, G. W., M. Kac, and P. Mazur, 1965, “Statistical mechanics of assemblies of coupled oscillators,” *J. Math. Phys. (N.Y.)* **6**, 504.
- Ford, G. W., J. T. Lewis, and R. F. O’Connell, 1985, “Quantum Oscillator in a Blackbody Radiation Field,” *Phys. Rev. Lett.* **55**, 2273.
- Fowler, R. H., 1936, *Statistical Mechanics*, 2nd ed. (Cambridge University Press, Cambridge, England).
- Francica, G., J. Goold, and F. Plastina, 2019, “Role of coherence in the nonequilibrium thermodynamics of quantum systems,” *Phys. Rev. E* **99**, 042105.
- Gardiner, C. W., 1985, 2nd ed. *Handbook of Stochastic Methods* (Springer-Verlag, Berlin).
- Gaspard, P., 2004, “Fluctuation theorem for nonequilibrium reactions,” *J. Chem. Phys.* **120**, 8898.
- Gelin, M. F., and M. Thoss, 2009, “Thermodynamics of a sub-ensemble of a canonical ensemble,” *Phys. Rev. E* **79**, 051121.
- Gibbs, J. W., 1902, *Elementary Principles in Statistical Mechanics* (Scribner’s Sons, New York).
- Goldstein, H., 2002, *Classical Mechanics* (Addison-Wesley, San Francisco).
- Goold, J., M. Huber, A. Riera, L. del Rio, and P. Skrzypczyk, 2016, “The role of quantum information in thermodynamics—A topical review,” *J. Phys. A* **49**, 143001.
- Goold, J., U. Poschinger, and K. Modi, 2014, “Measuring the heat exchange of a quantum process,” *Phys. Rev. E* **90**, 020101(R).
- Grabert, H., 1982, *Projection Operator Techniques in Nonequilibrium Statistical Mechanics*, Springer Tracts in Modern Physics (Springer-Verlag, Berlin).
- Grabert, H., P. Hänggi, and P. Talkner, 1980, “Microdynamics and nonlinear stochastic processes of gross variables,” *J. Stat. Phys.* **22**, 537.
- Grabert, H., P. Schramm, and G.-L. Ingold, 1988, “Quantum Brownian motion: The functional integral approach,” *Phys. Rep.* **168**, 115.
- Grabert, H., P. Talkner, and P. Hänggi, 1977, “Microdynamics and time-evolution of macroscopic non-Markovian systems,” *Z. Phys. B* **26**, 389.
- Grabert, H., U. Weiss, and P. Talkner, 1984, “Quantum theory of the damped harmonic oscillator,” *Z. Phys. B* **55**, 87.
- Hänggi, P., S. Hilbert, and J. Dunkel, 2016, “Meaning of temperature in different thermostistical ensembles,” *Phil. Trans. R. Soc. A* **374**, 20150039.
- Hänggi, P., and G.-L. Ingold, 2006, “Quantum Brownian motion and the third law of thermodynamics,” *Acta Phys. Pol. B* **37**, 1537.
- Hänggi, P., G.-L. Ingold, and P. Talkner, 2008, “Finite quantum dissipation: The challenge of obtaining specific heat,” *New J. Phys.* **10**, 115008.
- Hänggi, P., and P. Talkner, 2015, “The other QFT,” *Nat. Phys.* **11**, 108–110.
- Hänggi, P., P. Talkner, and M. Borkovec, 1990, “Reaction-rate theory—50 years after Kramers,” *Rev. Mod. Phys.* **62**, 251–341.
- Hänggi, P., and H. Thomas, 1982, “Stochastic processes: Time-evolution, symmetries and linear response,” *Phys. Rep.* **88**, 207.
- Harris, R. J., and G. M. Schütz, 2007, “Fluctuation theorems for stochastic dynamics,” *J. Stat. Mech.* P07020.
- Hatano, T., and S. Sasa, 2001, “Steady-State Thermodynamics of Langevin Systems,” *Phys. Rev. Lett.* **86**, 3463.
- Hertz, P., 1910a, “Über die mechanischen Grundlagen der Thermodynamik [part I],” *Ann. Phys. (Berlin)* **33**, 225.
- Hertz, P., 1910b, “Über die mechanischen Grundlagen der Thermodynamik [part II],” *Ann. Phys. (Berlin)* **33**, 537.
- Hilbert, S., P. Hänggi, and J. Dunkel, 2014, “Thermodynamic laws in isolated systems,” *Phys. Rev. E* **90**, 062116.
- Hoppenau, J., and A. Engel, 2013, “On the work distribution in quasi-static processes,” *J. Stat. Mech.* P06004.
- Hörhammer, C., and H. Büttner, 2008, “Information and entropy in quantum Brownian motion, thermodynamic entropy versus von Neumann entropy,” *J. Stat. Phys.* **133**, 1161.
- Hsiang, J.-T., C. H. Chou, Y. Subaşı, and B. L. Hu, 2018, “Quantum thermodynamics from the nonequilibrium dynamics of open systems: Energy, heat capacity and the third law,” *Phys. Rev. E* **97**, 012135.
- Hummer, G., and A. Szabo, 2001, “Free energy reconstruction from nonequilibrium single-molecule pulling experiments,” *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3658.
- Ingold, G. L., P. Hänggi, and P. Talkner, 2009, “Specific heat anomalies of open quantum systems,” *Phys. Rev. E* **79**, 061105.
- Ito, K., P. Talkner, B. P. Venkatesh, and G. Watanabe, 2019, “Generalized energy measurements and quantum work compatible with fluctuation theorems,” *Phys. Rev. A* **99**, 032117.
- Jarzynski, C., 1997a, “Nonequilibrium Equality for Free Energy Differences,” *Phys. Rev. Lett.* **78**, 2690.
- Jarzynski, C., 1997b, “Equilibrium free-energy differences from nonequilibrium measurements: A master equation approach,” *Phys. Rev. E* **56**, 5018.
- Jarzynski, C., 2004, “Nonequilibrium work theorem for a system strongly coupled to a thermal environment,” *J. Stat. Mech.* P09005.
- Jarzynski, C., 2007, “Comparison of far-from-equilibrium work relations,” *C.R. Phys.* **8**, 495.
- Jarzynski, C., 2011, “Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale,” *Annu. Rev. Condens. Matter Phys.* **2**, 329.
- Jarzynski, C., 2017, “Stochastic and Macroscopic Thermodynamics of Strongly Coupled Systems,” *Phys. Rev. X* **7**, 011008.
- Jaynes, E. T., and Cummings, 1963, “Comparison of quantum and semiclassical radiation theories with application to the beam maser,” *Proc. IEEE* **51**, 89.
- Jeon, E., P. Talkner, J. Yi, and Y. W. Kim, 2017, “Role of work in matter exchange between finite quantum systems,” *New J. Phys.* **19**, 093006.
- Kammerlander, P., and J. Anders, 2016, “Coherence and measurement in quantum thermodynamics,” *Sci. Rep.* **6**, 22174.
- Kantz, H., and T. Schreiber, 1997, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, England).
- Kawai, R., J. M. R. Parrondo, and C. Van den Broeck, 2007, “Dissipation: The Phase-Space Perspective,” *Phys. Rev. Lett.* **98**, 080602.
- Keizer, J., 1987, *Statistical Mechanics and Nonequilibrium Processes* (Springer-Verlag, Berlin).
- Khinchin, A. I., 1949, *Mathematical Foundations of Statistical Mechanics* (Dover Publications, New York).
- Kim, S., J. W. Kim, P. Talkner, and J. Yi, 2012, “Comparison of free-energy estimators and their dependence on dissipated work,” *Phys. Rev. E* **86**, 041130; **86**, 059903(E) (2012).
- Kirkwood, J. G., 1935, “Statistical mechanics of fluid mixtures,” *J. Chem. Phys.* **3**, 300.
- Kosloff, R., and Y. Rezek, 2017, “The quantum harmonic Otto cycle,” *Entropy* **19**, 136.
- Kubo, R., 1965, *Statistical Mechanics: An Advanced Course with Problems and Solutions* (North-Holland, Amsterdam), Chap. 1.6.
- Kurchan, J., 1998, “Fluctuation theorem for stochastic dynamics,” *J. Phys. A* **31**, 3719.

- Kurchan, J., 2000, “A quantum fluctuation theorem,” *arXiv: cond-mat/0007360*.
- Landauer, R., 1961, “Irreversibility and heat generation in the computing process,” *IBM J. Res. Dev.* **5**, 183.
- Lebowitz, J. L., and O. Penrose, 1973, “Modern ergodic theory,” *Phys. Today* **26**, No. 2, 23.
- Lechner, W., and C. Dellago, 2007, “On the efficiency of path sampling methods for the calculation of free energies from non-equilibrium simulations,” *J. Stat. Mech.* P04001.
- Lieb, E. H., and J. L. Lebowitz, 1972, “The constitution of matter: Existence of thermodynamics for systems composed of electrons and nuclei,” *Adv. Math.* **9**, 316.
- Liphardt, J., B. Onoa, S. B. Smith, I. Tinoco, and C. Bustamante, 2001, “Reversible unfolding of single RNA molecules by mechanical force,” *Science* **292**, 733–737.
- Lukacs, E., 1970, *Characteristic Functions*, 2nd ed. (Griffin, London).
- Lustig, R., 1994, “Statistical thermodynamics in the classical molecular dynamics ensemble. 1. Fundamentals,” *J. Chem. Phys.* **100**, 3048–3059.
- Magalinskii, V. B., 1959, “Dynamical model in the theory of Brownian motion,” *Zh. Eksp. Teor. Fiz.* **36**, 1942 [Sov. Phys. JETP **9**, 1381 (1959).]
- Marsh, C., 2013, “Introduction to continuous entropy,” [https://www.cmarsh.com/static/pdf/Charles\\_Marsh\\_Continuous\\_Entropy.pdf](https://www.cmarsh.com/static/pdf/Charles_Marsh_Continuous_Entropy.pdf).
- Merali, N., 2017, “Bending the rules: Thermodynamics might operate differently in the quantum realm. Experiments are starting to put that idea to the test,” *Nature (London)* **551**, 20.
- Messiah, A., 1962, *Quantum Mechanics*, Vol II (North-Holland, Amsterdam).
- Mikeska, H. J., and W. Pesch, 1977, “Boundary effects on static spin correlation-functions in isotropic  $X$ - $Y$  chain at zero temperature,” *Z. Phys. B* **26**, 351.
- Miller, H. J. D., and J. Anders, 2018, “Leggett-Garg inequality for quantum fluctuating work,” *Entropy* **20**, 200.
- Miller, H. J. D., M. Scandi, J. Anders, and M. Perarnau-Llobet, 2019, “Work Fluctuations in Slow Processes: Quantum Signatures and Optimal Control,” *Phys. Rev. Lett.* **123**, 230603.
- Miller, H. R., 2018, “Hamiltonian of mean force for strongly-coupled systems,” in *Thermodynamics in the Quantum Regime*, edited by F. Binder, L. A. Correa, C. Gogolin, J. Anders, and G. Adesso (Springer Nature, Cham, Switzerland), p. 531, Chap. 22.
- Münster, A., 1954, “Anwendung der  $\delta$ -Funktion auf die mikrokanonische Gesamtheit,” *Z. Phys.* **137**, 386–391.
- Nieuwenhuizen, T. M., and A. E. Allahverdyan, 2002, “Statistical thermodynamics of quantum Brownian motion: Construction of perpetual mobile of the second kind,” *Phys. Rev. E* **66**, 036102.
- Norton, J. D., 2013, “All shook up: Fluctuations, Maxwell’s demon and the thermodynamics of computation,” *Entropy* **15**, 4432.
- Onsager, L., 1933, “Theories of concentrated electrolytes,” *Chem. Rev.* **13**, 73.
- Perarnau-Llobet, M., E. Bäumer, K. V. Hovhannisyanyan, M. Huber, and A. Acin, 2017, “No-Go Theorem for the Characterization of Work Fluctuations in Coherent Quantum Systems,” *Phys. Rev. Lett.* **118**, 070601.
- Philbin, P., and J. Anders, 2016, “Thermal energies of classical and quantum damped oscillators coupled to reservoirs,” *J. Phys. A* **49**, 215303.
- Pohorille, A., C. Jarzynski, and C. Chipot, 2010, “Good practices in free energy calculations,” *J. Phys. Chem. B* **114**, 10235.
- Polkovnikov, A., K. Sengupta, A. Silva, and M. Vengalattore, 2011, “*Colloquium*: Nonequilibrium dynamics of closed interacting quantum systems,” *Rev. Mod. Phys.* **83**, 863–883.
- Rigol, M., and Srednicki, 2012, “Alternatives to Eigenstate Thermalization,” *Phys. Rev. Lett.* **108**, 110601.
- Risken, H., 1989, *The Fokker-Planck Equation*, 2nd ed. (Springer-Verlag, Berlin).
- Roux, B., 1995, “The calculation of the potential of mean force using computer simulations,” *Comput. Phys. Commun.* **91**, 275.
- Roux, B., and T. Simonson, 1999, “Implicit solvent models,” *Biophys. Chem.* **78**, 1.
- Ruelle, D., 1969, *Statistical Mechanics* (W. A. Benjamin, New York).
- Sampaio, R., S. Suomela, T. Ala-Nissila, J. Anders, and T. G. Philbin, 2018, “Thermodynamics of a subensemble of a canonical ensemble,” *Phys. Rev. A* **97**, 012131.
- Schatten, R., 1950, *The Theory of Cross-Spaces* (Princeton University Press, Princeton, NJ).
- Schlüter, A., 1948, “Zur Statistik klassischer Gesamtheiten,” *Z. Naturforsch. A* **3**, 350–360.
- Seifert, U., 2005, “Entropy Production along a Stochastic Trajectory and Integral Fluctuation Theorem,” *Phys. Rev. Lett.* **95**, 040602.
- Seifert, U., 2012, “Stochastic thermodynamics, fluctuation theorems and molecular machines,” *Rep. Prog. Phys.* **75**, 126001.
- Seifert, U., 2016, “First and Second Law of Thermodynamics at Strong Coupling,” *Phys. Rev. Lett.* **116**, 020601.
- Seifert, U., 2019, “From stochastic thermodynamics to thermodynamic inference,” *Annu. Rev. Condens. Matter Phys.* **10**, 171.
- Sekimoto, K., 1998, “Langevin equation and thermodynamics,” *Prog. Theor. Phys. Suppl.* **130**, 17.
- Sekimoto, K., 2010, *Stochastic Energetics*, 2nd ed., Lecture Notes in Physics Vol. 799 (Springer-Verlag, Berlin).
- Smith, A., Y. Lu, S. An, X. Zhang, Z. Gong, H. T. Quan, C. Jarzynski, and K. Kim, 2018, “Verification of the quantum non-equilibrium work relation in the presence of decoherence,” *New J. Phys.* **20**, 013008.
- Spreng, B., G.-L. Ingold, and U. Weiss, 2013, “Reentrant classicality of a damped system,” *Europhys. Lett.* **103**, 60007.
- Srednicki, M., 1994, “Chaos and quantum thermalization,” *Phys. Rev. E* **50**, 888.
- Strasberg, P., and M. Esposito, 2017, “Stochastic thermodynamics in the strong coupling regime: An unambiguous approach based on coarse graining,” *Phys. Rev. E* **95**, 062101.
- Strasberg, P., and M. Esposito, 2020, “On the measurability of nonequilibrium thermodynamics in terms of the Hamiltonian of mean force,” *arXiv:2001.08917*.
- Strasberg, P., G. Schaller, T. Brandes, and M. Esposito, 2017, “Quantum and Information Thermodynamics: A Unifying Framework Based on Repeated Interactions,” *Phys. Rev. X* **7**, 021003.
- Stratonovich, R. L., 1963, *Topics in the Theory of Random Noise*, Vol. I (Gordon and Breach, New York).
- Talkner, P., 1981, “Gauss Markov process of a quantum oscillator,” *Z. Phys. B* **41**, 365.
- Talkner, P., P. S. Burada, and P. Hänggi, 2008, “Statistics of work performed on a forced quantum oscillator,” *Phys. Rev. E* **78**, 011115.
- Talkner, P., M. Campisi, and P. Hänggi, 2009, “Fluctuation theorems in driven open quantum systems,” *J. Stat. Mech.* P02025.
- Talkner, P., and P. Hänggi, 2007, “The Tasaki-Crooks quantum fluctuation theorem,” *J. Phys. A* **40**, F569.
- Talkner, P., and P. Hänggi, 2016a, “Aspects of quantum work,” *Phys. Rev. E* **93**, 022131.
- Talkner, P., and P. Hänggi, 2016b, “Open system trajectories specify fluctuating work but not heat,” *Phys. Rev. E* **94**, 022143.
- Talkner, P., and P. Hänggi, 2020, “Comment on ‘Measurability of nonequilibrium thermodynamics in terms of the Hamiltonian of mean force,’” *arXiv:2006.16592*

- Talkner, P., P. Hänggi, and M. Morillo, 2008, “Microcanonical quantum fluctuation theorems,” *Phys. Rev. E* **77**, 051131.
- Talkner, P., E. Lutz, and P. Hänggi, 2007, “Fluctuation theorems: Work is not an observable,” *Phys. Rev. E* **75**, 050102.
- Talkner, P., M. Morillo, J. Yi, and P. Hänggi, 2013, “Statistics of work and fluctuation theorems for microcanonical initial states,” *New J. Phys.* **15**, 095001.
- Tasaki, H., 2000, “Jarzynski relations for quantum systems and some applications,” [arXiv:cond-mat/0009244](https://arxiv.org/abs/cond-mat/0009244).
- Thirring, W., H. Narnhofer, and H. A. Posch, 2003, “Negative Specific Heat, the Thermodynamic Limit, and Ergodicity,” *Phys. Rev. Lett.* **91**, 130601.
- Touchette, H., 2009, “The large deviation approach to statistical mechanics,” *Phys. Rep.* **478**, 1.
- Touchette, H., 2011, “Ensemble equivalence for general many-body systems,” *Europhys. Lett.* **96**, 50010.
- Ullersma, P., 1966, “An exactly solvable model for Brownian motion. I. Derivation of Langevin equation,” *Physica (Utrecht)* **32**, 27.
- Valente, D. M., and A. O. Caldeira, 2010, “Thermal equilibrium of two quantum Brownian particles,” *Phys. Rev. A* **81**, 012117.
- Van den Broeck, C., and M. Esposito, 2015, “Ensemble and trajectory thermodynamics: A brief introduction,” *Physica (Amsterdam)* **418A**, 6.
- Van Hove, L., 1957, “The approach to equilibrium in quantum statistics—A perturbation treatment to general order,” *Physica (Utrecht)* **23**, 441–480.
- van Kampen, N. G., 2007, *Stochastic Processes in Physics and Chemistry*, 3rd ed. (Elsevier, Amsterdam).
- Venkatesh, B. P., G. Watanabe, and P. Talkner, 2015, “Quantum fluctuation theorems and power measurements,” *New J. Phys.* **17**, 075018.
- Vinjanampathy, S., and J. Anders, 2016, “Quantum thermodynamics,” *Contemp. Phys.* **57**, 545.
- von Laue, M., 1950, *History of Physics* (Academic Press, New York).
- von Neumann, J., 1929, “Beweis des Ergodensatzes und des *H*-Theorems in der neuen Mechanik,” *Z. Phys.* **57**, 30–70 [“Proof of the ergodic theorem and the *H*-theorem in quantum mechanics,” *Eur. Phys. J. H.* **35**, 201–237 (2010)].
- von Neumann, J., 1955, *Mathematical Foundation of Quantum Mechanics* (Princeton University Press, Princeton).
- Watanabe, G., B. P. Venkatesh, and P. Talkner, 2014, “Generalized energy measurements and modified transient quantum fluctuation theorems,” *Phys. Rev. E* **89**, 052116.
- Watanabe, G., B. P. Venkatesh, P. Talkner, M. Campisi, and P. Hänggi, 2014, “Quantum fluctuation theorems and generalized measurements during the force protocol,” *Phys. Rev. E* **89**, 032114.
- Wehrl, A., 1978, “General properties of entropy,” *Rev. Mod. Phys.* **50**, 221–260.
- Weiss, U., 2008, *Quantum Dissipative Systems*, 3rd ed. (World Scientific, Singapore).
- Yang, D., A. Grankin, L. Sieberer, D. V. Vasilyev, and P. Zoller, 2020, “Quantum nondemolition measurement of a many-body Hamiltonian,” *Nat. Commun.* **11**, 775.
- Yi, J., Y. W. Kim, and P. Talkner, 2012, “Work fluctuations for Bose particles in grand canonical initial states,” *Phys. Rev. E* **85**, 051107.
- Yi, J., P. Talkner, and M. Campisi, 2011, “Nonequilibrium work statistics of an Aharonov-Bohm flux,” *Phys. Rev. E* **84**, 011138.
- Zheng, Y. J., P. Hänggi, and D. Poletti, 2016, “Occurrence of discontinuities in the performance of finite-time quantum Otto cycles,” *Phys. Rev. E* **94**, 012137.
- Zwanzig, R., 1961, “Memory effects in reversible thermodynamics,” *Phys. Rev.* **124**, 983.
- Zwanzig, R., 1973, “Nonlinear generalized Langevin equations,” *J. Stat. Phys.* **9**, 215.
- Zwanzig, R., 2001, *Nonequilibrium Statistical Mechanics* (Oxford University Press, Oxford).

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

Peter Talkner and Peter Hänggi

 (published 29 July 2022)

DOI: [10.1103/RevModPhys.94.039901](https://doi.org/10.1103/RevModPhys.94.039901)

In the text preceding Eq. (115) in the Colloquium, the transition probability  $p_{\Lambda}(m', \alpha' m, \alpha)$  between eigenspaces of the Hamiltonians of the system and its environment with energies  $E_{m'}(\tau)$ ,  $\epsilon_{\alpha'}$  and  $E_m(0)$ ,  $\epsilon_{\alpha}$  caused by a force protocol of duration  $\tau$  acting only on the system is correct up to a mistyped index and a missing bath projection operator  $Q_{\alpha}$ ,  $p_{\Lambda}(m', \alpha' m, \alpha) = \text{Tr}_{\text{tot}} \Pi_{m'}(\tau) Q_{\alpha'} U_{\Lambda} \Pi_m(0) Q_{\alpha} U_{\Lambda}^{\dagger} / \text{Tr}_{\text{tot}} \Pi_m Q_{\alpha}$ , provided that the initial eigenspaces with the system and bath energies  $E_m$  and  $\epsilon_{\alpha}$  are uniformly populated, i.e., if  $\Pi_m Q_{\alpha} \rho(0) Q_{\alpha} \Pi_m \propto \Pi_m Q_{\alpha}$  holds. In particular, the previously mentioned form holds for nondegenerate energies. In general, the transition probability is given by [see also [Talkner, Campisi, and Hänggi \(2009\)](#)]

$$p_{\Lambda}(m', \alpha' m, \alpha) = \text{Tr}_{\text{tot}} \Pi_{m'}(\tau) Q_{\alpha'} U_{\Lambda} \rho_{m,\alpha} U_{\Lambda}^{\dagger} \quad (1)$$

where

$$\rho_{m,\alpha} = \Pi_m Q_{\alpha} \rho(0) Q_{\alpha} \Pi_m / \text{Tr}_{\text{tot}} \Pi_m Q_{\alpha} \rho(0) \quad (2)$$

specifies the initial state restricted to the subspace with energies  $E_m(0)$  and  $\epsilon_{\alpha}$ .

The preprint cited in the Colloquium as [Talkner and Hänggi \(2020\)](#) has since been published; see [Talkner and Hänggi \(2020\)](#).

### REFERENCES

- Talkner, P., M. Campisi, and P. Hänggi, 2009, “Fluctuation theorems in driven open quantum systems,” *J. Stat. Mech.* P02025.  
Talkner, P., and P. Hänggi, 2020, “Comment on ‘Measurability of nonequilibrium thermodynamics in terms of the Hamiltonian of mean force,’” *Phys. Rev. E* **102**, 066101.

**Comment on “Measurability of nonequilibrium thermodynamics in terms of the Hamiltonian of mean force”**

Peter Talkner <sup>1</sup> and Peter Hänggi <sup>1,2,3</sup>

<sup>1</sup>*Department of Physics, University of Augsburg, D-86135 Augsburg, Germany*

<sup>2</sup>*Nanosystems Initiative Munich, D-80799 München, Germany*

<sup>3</sup>*Department of Physics, National University of Singapore, Singapore 117546*



(Received 29 June 2020; accepted 23 November 2020; published 21 December 2020)

In a recent paper [Phys. Rev. E **101**, 050101(R) (2020)] an attempt is presented to formulate the nonequilibrium thermodynamics of an open system in terms of the Hamiltonian of mean force. The purpose of the present comment is to clarify severe restrictions of this approach and to stress that recently noted ambiguities [Phys. Rev. E **94**, 022143 (2016)] of fluctuating thermodynamic potentials cannot be removed in the suggested way.

DOI: 10.1103/PhysRevE.102.066101

The Hamiltonian of mean force [1], such as its classical precursor, the potential of mean force [2,3], encompasses the complete information on the equilibrium thermodynamical behavior of an open system. It is defined in terms of the bath-renormalized Boltzmann factor at the inverse temperature  $\beta$ , given by the expression,

$$e^{-\beta H^*(\beta, \mu)} = Z_B^{-1}(\beta, \mu) \text{Tr}_B e^{-\beta H_{\text{tot}}(\mu)}, \quad (1)$$

where  $H^*(\beta, \mu)$  and  $H_{\text{tot}}(\mu) = H_S(\mu) + H_{SB}(\mu) + H_B(\mu)$  denote the Hamiltonian of mean force, the Hamiltonian of the system  $H_S(\mu)$ , and the environment  $H_B(\mu)$ , including the mutual interaction  $H_{SB}(\mu)$ , respectively. With  $\mu$ , all externally controllable parameters, including those parameters  $\lambda$ , exclusively acting on the considered open system, but also including all other globally acting parameters, such as electric and magnetic fields, are specified. Furthermore,  $\text{Tr}_B$  indicates the partial trace over the environmental Hilbert space for quantum systems and, for a classical system, the integral over the environmental phase space with respect to a properly permutation-symmetry-adapted dimensionless volume element. With the normalization by the partition function of the environment,  $Z_B(\beta, \mu) = \text{Tr}_B e^{-\beta H_B(\mu)}$ , the Hamiltonian of mean force agrees with the bare system Hamiltonian for isolated systems; for further details see Refs. [4,5]. With the partition function for the Hamiltonian of mean force  $Z_S(\beta, \mu) = \text{Tr}_S e^{-\beta H^*(\beta, \mu)} = Z_{\text{tot}}(\beta, \mu)/Z_B(\beta, \mu)$  and the resulting Helmholtz free energy  $F_S(\beta, \mu) = -\beta^{-1} \ln Z_S(\beta, \mu)$ , the equilibrium thermodynamics of the open system becomes accessible. Likewise, for a quantum system, the density matrix and, for a classical system, the respective phase space probability density function (pdf) are specified in terms of the Hamiltonian of mean force according to

$$\rho(\beta, \mu) = Z_S^{-1}(\beta, \mu) e^{-\beta H^*(\beta, \mu)}. \quad (2)$$

On the other hand, from the knowledge of  $\rho(\beta, \mu)$ , which can, in principle, be inferred from a system-intrinsic point of view, it is not possible to separate  $\rho(\beta, \mu)$  into the Hamiltonian of mean force and the system’s partition function in

conflict with a central dogma of stochastic thermodynamics [6–8] which postulates that all necessary information can be obtained from observations of the system without recourse to data from the environment. As a resort, Strasberg and Esposito argue in Ref. [8] that only those differences of free energies or, equivalently, those ratios of the system’s partition functions are relevant for the description of thermodynamic processes that are taken at different values of the direct system control parameters  $\lambda$ . With this assumption, the large class of thermodynamic processes accompanied by changes in global parameters as well as of temperature and pressure is excluded and most relevant thermodynamic quantities, such as specific heat, isobaric thermal expansion, magnetization, and polarization, magnetic, and electric susceptibilities, and compression factors are not accessible within such an approach. Hence, even though the description of special processes may not require extra information about the environment, in general, a system intrinsic description, i.e., one that is exclusively based on the observation of the system, be it by means of quantum tomography or of a monitoring of the stochastic trajectories of a classical open system, is not possible [4,5]. In contrast to Strasberg and Esposito [8] being concerned with the nonequilibrium dynamics of a special class of problems, our main aim in Refs. [4,5] is the characterization of the thermal equilibrium of open systems in its dependence on *all* relevant parameters. In this context, the modification of the Hamiltonian of mean force suggested by Strasberg and Esposito [8] cannot be qualified as an addition of an “irrelevant constant” that would have “no thermodynamic consequences”; the quotations are taken from Ref. [8]. For general thermodynamic processes it would rather lead to erroneous conclusions. Even in the restricted situation of isothermal processes at *constant* global parameters it is not sufficient to know the reduced density matrix or the pdf of the open system to infer changes in thermodynamic quantities. Additionally, process-specific relations, such as the Jarzynski equality or second-law-like relations must be imposed. This further narrows the predictive power of the method suggested by Ref. [8]. Moreover, the

inference of free energy differences on the basis of the Jarzynski equality often requires an unrealistically large amount of data [9,10]. The same kind of problem must be expected also for the other methods, in particular, when the tomography of time-dependent states is required to estimate, e.g., an effective system Hamiltonian on the basis of Eq. (23) in Ref. [8].

As explained in Refs. [4,5], the concept of fluctuating thermodynamic potentials suffers already in thermal equilibrium from ambiguities that can be subsumed as the set of all functions with a vanishing equilibrium average value with respect to the respective Gibbs state of the total system. For transient and other nonequilibrium processes, the respective set of functions is characterized by a vanishing average with respect to the actual time-dependent state of the system. It, hence, changes with time in dependence of the initial pdf of the system as it follows from Eqs. (53) and (54) in combination with Eqs. (50) and (51) of Ref. [4], but it does *not* collapse to an empty set as Strasberg and Esposito wrongly conclude in Footnote [60] of Ref. [8]. We further note that even the specification of a particular fluctuating free energy as performed in Eq. (5) of Ref. [11] does still leave the fluctuating internal energy and entropy largely unspecified as can be seen from the Eqs. (107) and (108) in Ref. [5]. In the particular case of an equilibrium system the supposedly fluctuating free energy defined in (5) of Ref. [11] yields the nonfluctuating equilibrium free energy [4].

Furthermore, we would like to clarify two misleading literal citations in Ref. [8] which are taken out of their original context. Our statement in Ref. [5] that “...presents in practice

an impossible task” does not refer to the Hamiltonian of mean force as insinuated by Strasberg and Esposito [8] but to the reconstruction of the total system’s Hamiltonian solely based on open system’s trajectories. Our observation that the first law of thermodynamics for quantum open systems interacting with their environments at a finite strength is doubtful is based on the fact that then the respective observables that determine work and heat do not commute. Hence, their simultaneous measurement is excluded by the laws of quantum mechanics. The condition that measurements “need to be error free”, cited from Ref. [5], refers to one of the mathematical properties that generalized energy measurements must satisfy in order to yield the Crooks and the Jarzynski fluctuation relations [12,13] but has not been made in the context of the first law.

Finally, we would like to stress that, by its very definition, the Hamiltonian of mean force describes thermodynamic *equilibrium*. Expressions, such as the nonequilibrium free energy in Eq. (5) of Ref. [8] or the corresponding fluctuating nonequilibrium free energy in Eq. (5) of Ref. [11], based on the Hamiltonian of mean force and the system’s pdf at time  $t$  and at the position of a random trajectory, are mere postulates without a deeper rationale. The latter object, resulting in the described way from the pdf, is by construction a function of the starting point of the considered random trajectory and a functional of the random force having acted up to time  $t$ . This very construct does not follow as a transformation of a pdf according to the proper rules of probability theory. It, therefore, is not normalized with respect to the starting point and, hence, has no obvious probabilistic meaning, see the Supplemental Material [14].

- 
- [1] M. Campisi, P. Talkner, and P. Hänggi, Fluctuation Theorem for Arbitrary Open Quantum Systems, *Phys. Rev. Lett.* **102**, 210401 (2009).
- [2] L. Onsager, Theories of concentrated electrolytes, *Chem. Rev.* **13**, 73 (1933).
- [3] G. Kirkwood, Statistical mechanics of fluid mixtures, *J. Chem. Phys.* **3**, 300 (1935).
- [4] P. Talkner and P. Hänggi, Open system trajectories specify fluctuating work but not heat, *Phys. Rev. E* **94**, 022143 (2016).
- [5] P. Talkner and P. Hänggi, Colloquium: Statistical mechanics and thermodynamics at strong coupling: Quantum and classical, *Rev. Mod. Phys.* **92**, 041002 (2020).
- [6] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, *Rep. Prog. Phys.* **75**, 126001 (2012).
- [7] U. Seifert, From stochastic thermodynamics to thermodynamic inference, *Annu. Rev. Condens. Matter Phys.* **10**, 171 (2019).
- [8] P. Strasberg and M. Esposito, Measurability of nonequilibrium thermodynamics in terms of the Hamiltonian of mean force, *Phys. Rev. E* **101**, 050101(R) (2020).
- [9] S. Kim, Y. W. Kim, P. Talkner, and J. Yi, Comparison of free-energy estimators and their dependence on dissipated work, *Phys. Rev. E* **86**, 041130 (2012).
- [10] J. W. Deng, A. M. Tan, P. Hänggi, and J. B. Gong, Merits and qualms of work fluctuations theorems in classical fluctuation theorems, *Phys. Rev. E* **95**, 012106 (2017).
- [11] See Supplemental Material of Ref. [8].
- [12] B. P. Venkatesh, G. Watanabe, and P. Talkner, Transient quantum fluctuation theorems and generalized measurements, *New J. Phys.* **16**, 015032 (2014).
- [13] K. Ito, P. Talkner, B. P. Venkatesh, and G. Watanabe, Generalized energy measurements and quantum work compatible with fluctuation theorems, *Phys. Rev. A* **99**, 032117 (2019).
- [14] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevE.102.066101> for the discussion of a particular exemplary case, which includes Refs. [15–17].
- [15] U. Seifert, Entropy Production Along a Stochastic Trajectory and an Integral Fluctuation Theorem, *Phys. Rev. Lett.* **95**, 040602 (2005).
- [16] M. C. Wang and G. E. Uhlenbeck, On the theory of the Brownian motion, *Rev. Mod. Phys.* **17**, 323 (1945).
- [17] D. Ludwig, Persistence of dynamical systems under random perturbations, *SIAM Rev.* **17**, 605 (1975).

— Supplemental Material —  
**Comment on “Measurability of nonequilibrium thermodynamics in terms of the Hamiltonian of mean force**

Peter Talkner<sup>1,\*</sup> and Peter Hänggi<sup>1,2,3,†</sup>

<sup>1</sup>*Department of Physics, University of Augsburg, D 86135 Augsburg, Germany*

<sup>2</sup>*Nanosystems Initiative Munich, D-80799 München, Germany*

<sup>3</sup>*Department of Physics, National University of Singapore, Singapore 117546*

**Contents**

<b>I.</b>	<b>“fluctuating probability density function” of stochastic thermodynamics</b>	1
<b>II.</b>	<b>Brownian harmonic oscillator</b>	1
	A. Langevin equation and the probability density function	1
	B. The “Fluctuating probability density function”	2
	<b>References</b>	3

**I. THE “FLUCTUATING PROBABILITY DENSITY FUNCTION” OF STOCHASTIC THERMODYNAMICS**

The theory of stochastic thermodynamics is based on the hermaphroditic notion of a fluctuating probability density function (fpdf) whose functional form results from a *proper* probability density function (pdf) in which the state space variable is substituted by a particular realization of the considered process [1–3]. In the present supplemental material we would like to illustrate this construct, combining aspects of states and observables as well as forward and backward dynamics, with the example of Markovian diffusion processes. Finally we shall specialize to the case of a Brownian oscillator. If we denote a point in the state space of the considered process as  $\mathbf{z}$ , the pdf, which is a solution of the Fokker-Planck equation, by  $\rho(\mathbf{z}, t)$  and a solution of the Langevin equation by  $\mathbf{Z}(\mathbf{z}, t)$  with  $\mathbf{Z}(\mathbf{z}, 0) = \mathbf{z}$ , the fpdf  $\varphi(\mathbf{z}, t)$  is defined according to [1–3] as

$$\varphi(\mathbf{z}, t) = \rho(\mathbf{Z}(\mathbf{z}, t), t), \quad (1)$$

which is a function of the starting point of the considered realization and no longer of its end point. Trivially, at the initial time  $t = 0$  the fpdf and pdf agree with each other. In the sequel we demonstrate by means of the example of a Brownian oscillator that the fpdf does *not* stay normalized, in general, **as its very construction does not conform with the transformation rules of probability densities. An exception presents a Hamiltonian dynamics, as shown below.**

**II. BROWNIAN HARMONIC OSCILLATOR**

**A. Langevin equation and the probability density function**

The process of a damped harmonic oscillator of mass  $m$  and frequency  $\omega$  under the influence of a Gaussian white random force  $\xi(t)$  is described by the Langevin equation [4]

$$\begin{aligned} \dot{q}(t) &= \frac{p(t)}{m} \\ \dot{p}(t) &= -\gamma p(t) - m\omega^2 q(t) + \xi(t), \end{aligned} \quad (2)$$

---

\*Electronic address: peter.talkner@physik.uni-augsburg.de

†Electronic address: hanggi@physik.uni-augsburg.de

where  $q(t)$  and  $p(t)$  are the position and momentum, respectively, of the oscillator at the time  $t$ , and  $\gamma$  denotes the friction constant. The average of the random force  $\xi(t)$  vanishes and its auto-correlation function is given by

$$\langle \xi(t)\xi(s) \rangle = 2m\gamma k_B T \delta(t-s), \quad (3)$$

depending on the temperature  $T$  of the bath, causing the frictional and fluctuating forces. Here,  $k_B$  denotes the Boltzmann constant. The vector  $\mathbf{Z}(\mathbf{z}, t) = (Q(\mathbf{z}, t), P(\mathbf{z}, t))$  of the solutions  $Q(\mathbf{z}, t)$  and  $P(\mathbf{z}, t)$  of Eq. 2 starting at  $\mathbf{Z}(\mathbf{z}, 0) = \mathbf{z} = (q, p)$  can be written as

$$\mathbf{Z}(\mathbf{z}, t) = \mathbf{Z}_h(\mathbf{z}, t) + \mathbf{Z}_i(t), \quad (4)$$

where

$$\begin{aligned} \mathbf{Z}_h(\mathbf{z}, t) &= e^{Rt} \mathbf{z} \\ \mathbf{Z}_i(t) &= \int_0^t ds e^{R(t-s)} \boldsymbol{\xi}(s). \end{aligned} \quad (5)$$

Here,  $R$  is the matrix of the coefficients of  $q(t)$  and  $p(t)$  on the right hand side of Eq. 2, hence reading

$$R = \begin{pmatrix} 0 & 1/m \\ -m\omega^2 & -\gamma \end{pmatrix}. \quad (6)$$

Accordingly, the exponentiated matrix becomes

$$e^{Rt} = e^{-\gamma t/2} \begin{pmatrix} \cos \omega_\gamma t + \frac{\gamma}{2\omega_\gamma} \sin \omega_\gamma t & \frac{1}{m\omega_\gamma} \sin \omega_\gamma t \\ -\frac{m\omega^2}{\omega_\gamma} \sin \omega_\gamma t & \cos \omega_\gamma t - \frac{\gamma}{2\omega_\gamma} \sin \omega_\gamma t \end{pmatrix}, \quad (7)$$

where  $\omega_\gamma = (\omega^2 - \gamma^2/4)^{1/2}$  is the effective frequency of the oscillator. For the sake of simplicity we assume  $\omega > \gamma/2$ . Finally,  $\boldsymbol{\xi}(t) = (0, \xi(t))$  denotes the vectorial random force.

In order to construct the fpdf we need the time-dependent pdf  $\rho(\mathbf{z}, t)$ . Due to the linearity of the process defined by the Langevin equation (2), an initially Gaussian pdf stays Gaussian for all later times  $t$ , taking the form

$$\rho(q, p, t) = (2\pi M(t))^{-1/2} e^{-\frac{1}{2}(\mathbf{z} - \langle \mathbf{Z}(t) \rangle)^{tr} \cdot M^{-1}(t) \cdot (\mathbf{z} - \langle \mathbf{Z}(t) \rangle)^T} \quad (8)$$

with the superscript  $tr$  indicating the transposed respective vector or matrix. According to the Eqs. (4) and (5), the average  $\langle \mathbf{Z}(t) \rangle$  becomes

$$\langle \mathbf{Z}(t) \rangle = e^{Rt} \langle \mathbf{z} \rangle, \quad (9)$$

where  $\langle \mathbf{z} \rangle$  denotes the average of  $\mathbf{z}$  with respect to the initial distribution. The time-dependence of the covariance matrix  $M(t)$  of the vector  $\mathbf{Z}$  follows from the following equation of motion [5]

$$\dot{M}(t) = M(t)R^{tr} + RM(t) + 2D \quad (10)$$

with  $D$  denoting the diffusion matrix reading

$$D = \begin{pmatrix} 0 & 0 \\ 0 & m\gamma k_B T \end{pmatrix} \quad (11)$$

yielding for the time-dependent autocorrelation matrix the expression

$$M(t) = e^{Rt} \left[ M(0) + 2 \int_0^t ds e^{-Rs} D e^{-R^{tr}s} \right] e^{R^{tr}t} \quad (12)$$

## B. The ‘‘Fluctuating probability density function’’

Combining the Eqs. (1) and (8) with (4,5) and (12) one obtains

$$\varphi(\mathbf{z}, t) = (2\pi M(t))^{-1/2} e^{-1/2(\delta z - \mathbf{v}(t))^{tr} Q^{-1}(t) (\delta z - \mathbf{v}(t))} \quad (13)$$

where  $\delta\mathbf{z} = \mathbf{z} - \langle\mathbf{z}\rangle_0$  denotes the fluctuations of the phase space points  $\mathbf{z}$  in the initial ensemble specified by  $\rho(\mathbf{z}, 0)$ ,  $\mathbf{v}(t) = \int_0^t ds e^{-Rs} \boldsymbol{\xi}(s)$  and  $Q(t) = M(0) + 2 \int_0^t ds e^{-Rs} D e^{-R^T r s}$ . The integral of the fpdf extended over all initial phase space points then results in

$$\begin{aligned} \int d\mathbf{z} \varphi(\mathbf{z}, t) &= \left[ \frac{\det Q(t)}{\det M(t)} \right]^{1/2} \\ &= \det e^{-Rt/2} = e^{\gamma t/2} \end{aligned} \quad (14)$$

where we used Eq. (7). Therefore, the fluctuating probability density does in general not stay normalized to one and hence cannot be interpreted as a probability density.

In the absence of friction, for purely Hamiltonian dynamics, the substitution of a time evolved trajectory in the probability density leads back to the initial pdf as can be seen by writing the time-evolved pdf  $\rho(\mathbf{z}, t)$  in terms of the initial pdf  $\rho(\mathbf{z}, 0)$  according to the general transformation rules of pdfs as

$$\rho(\mathbf{z}, t) = \int d\mathbf{z}_0 \delta(\mathbf{z} - \mathbf{Z}(\mathbf{z}_0, t)) \rho(\mathbf{z}_0, 0) = \rho(\mathbf{Z}^{-1}(\mathbf{z}, t), 0), \quad (15)$$

with the Hamiltonian trajectory  $\mathbf{Z}(\mathbf{z}, t)$ , mapping the initial to the final phase space point and its uniquely defined inverse  $\mathbf{Z}^{-1}(\mathbf{z}, t)$  acting oppositely. Hence, the replacement of  $\mathbf{z}$  by the Hamiltonian trajectory  $\mathbf{Z}(\mathbf{z}, t)$  in the time-dependent pdf leads back to the initial pdf according to

$$\varphi(\mathbf{z}, t) = \rho(\mathbf{Z}(\mathbf{z}, t), t) = \rho(\mathbf{z}, 0). \quad (16)$$

Finally, we note that the time dependence of the integral in Eq. (14) is due to the dissipation, but not due to the randomness of the Brownian oscillator dynamics.

- 
- [1] U. Seifert, *Entropy production along a stochastic trajectory and an integral fluctuation theorem*, Phys. Rev. Lett. **95**, 040602 (2005).  
[2] U. Seifert, *Stochastic thermodynamics, fluctuation theorems and molecular machines*, Rep. Prog. Phys. **75**, 126001 (2012).  
[3] P. Strasberg and M. Esposito, *Measurability of nonequilibrium thermodynamics in terms of the Hamiltonian of mean force*, Phys. Rev. E **101**, 050101 (2020).  
[4] M.C. Wang, G.E. Uhlenbeck, *On the theory of the Brownian motion*, Rev. Mod. Phys. **17**, 323 (1945).  
[5] D. Ludwig, *Persistence of Dynamical Systems under Random Perturbations*, SIAM Review, **17**, 605 (1975).