Non-Markovian quantum thermodynamics: Laws and fluctuation theorems

Robert S. Whitney
Laboratoire de Physique et Modélisation des Milieux Condensés (UMR 5493), Université Grenoble Alpes and CNRS,
Maison des Magistères, BP 166, 38042 Grenoble, France

(Received 6 November 2016; revised manuscript received 15 June 2018; published 7 August 2018)

This work brings together thermodynamics and nonequilibrium quantum theory, by showing that a real-time diagrammatic technique on the Keldysh contour is an equivalent of stochastic thermodynamics for non-Markovian quantum machines (heat engines, refrigerators, etc.). Symmetries are found between quantum trajectories and their time reverses on the Keldysh contour, for any interacting quantum system coupled to ideal reservoirs of electrons, phonons, or photons. These lead to quantum fluctuation theorems the same as the well-known classical ones (Jarzynski and Crooks equalities, integral fluctuation theorem, etc.), whether the system's dynamics are Markovian or not. Some of these are also shown to hold for nonfactorizable initial states. The sequential tunneling approximation and the cotunneling approximation are both shown to respect the symmetries that ensure the fluctuation theorems. For all initial states, energy conservation ensures that the first law of thermodynamics holds on average, while the above symmetries ensure that the second law of thermodynamics holds on average, even if fluctuations violate it.

DOI: 10.1103/PhysRevB.98.085415

I. INTRODUCTION

The laws of thermodynamics were derived for macroscopic machines, where entropy-reducing fluctuations (e.g., a gas spontaneously drifting into one corner of its container) are so rare that they have been referred to as *thermodynamic miracles* [1]. In microscopic systems on short timescales, these "miracles" are rather common, and we now know they obey fluctuation theorems [2–6]. There is a unifying theory of such theorems in classical systems called *stochastic thermodynamics* [7,8], reviewed in Refs. [9–12]. It gives the Jarzynski [13], Evans-Searles [14] and Crooks [15,16] equalities in the relevant limits. It was used to show [7,8] that *any* classical system with Markovian dynamics obeys

$$\langle e^{-\Delta S_{\mathrm{tot}}}\rangle = 1, \tag{1}$$

where ΔS_{tot} is the total entropy change of the system and reservoirs [17] and the average is over all possible thermal fluctuations [18]. This has become known as the *integral fluctuation theorem* [7–10], even if a similar identity had appeared under the name *nonequilibrium partition identity* earlier [19–21]. Equation (1) tells us that the second law of thermodynamics is obeyed on average, $\langle \Delta S_{\mathrm{tot}}\rangle \geqslant 0$. Yet Eq. (1) also tells us that fluctuations with $\Delta S_{\mathrm{tot}} < 0$ *must* occur (even if rarely); otherwise $\langle e^{-\Delta S_{\mathrm{tot}}}\rangle$ would be less than one.

At the same time, there is great interest in the thermodynamics of nanoscale machines, particularly those which convert heat into electricity, or use electricity to perform refrigeration. Such machines are definitely not macroscopic, so we can expect them to exhibit fluctuations similar to those described above. However most of them also exhibit quantum effects that are not captured by classical theory of stochastic thermodynamics. Many operate in the steady state, such as the quantum-dot heat engines experimentally realized in Refs. [22–24], or other mesoscopic systems which exhibit thermoelectric effects [11,25], while others involve pumping cycles [26]. The general case of such a machine is sketched in Fig. 1(a).

This work shows that a diagrammatic technique on the Keldysh contour—real-time transport theory [27–30]—provides an equivalent of stochastic thermodynamics for any quantum system coupled to reservoirs [Fig. 1(a)] whether that system's dynamics are Markovian or not. It makes the connection between the contribution of a double trajectory, γ, on the Keldysh contour and the contribution of its time reverse, $\overline{\gamma}$ [Fig. 5(a)]. This is enough to show that such systems respect the same fluctuation theorems as classical Markovian systems, and so obey the second law of thermodynamics on average. For the second law, our proof goes beyond those for Markovian quantum systems [31], those for systems with mean-field interactions [32,33], and Keldysh treatments for noninteracting systems (quadratic Hamiltonians) [34–36] or adiabatic driving [37] based on the Keldysh techniques reviewed in Ref. [38]. This connection between fluctuation theorems [4–6] and a nonequilibrium quantum theory for transport through interacting systems [27–30,39–43] provides a powerful tool for modeling energy production and refrigeration at the nanoscale. In this context, significant currents and power outputs require significant system-reservoir coupling. However, only systems in the weak-coupling limit have Markovian dynamics [44,45]. Thus there is great interest in improving the power output of experimental setups like the quantum dot heat engines in Refs. [22–24] by taking them to stronger coupling, where their dynamics will be non-Markovian systems.

Previous proofs of fluctuation theorems in non-Markovian quantum systems exist [4] but rely on treating the system and reservoirs together as a single isolated quantum system. This is elegant, but not amenable to calculating a given machine's power or efficiency, except in the rare cases where the full

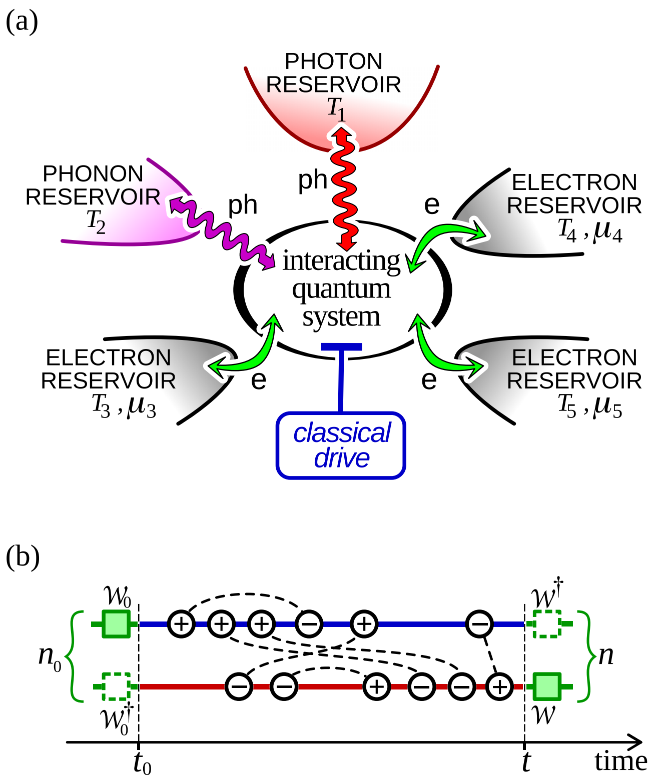

FIG. 1. (a) This work considers a quantum system coupled to any number of electron reservoirs with chemical potentials and temperatures $\{\mu_\alpha, T_\alpha\}$, and photon or phonon reservoirs at temperatures $\{T_\alpha\}$. (b) A typical double Keldysh trajectory, γ, in which the horizontal lines represent the evolution of the system state, while the dashed lines indicate transitions within the system due to the coupling to one of the reservoirs. (a) Typical set-up considered here (b) Typical trajectory on Keldysh contour.

Hamiltonian (system plus reservoirs) is exactly soluble. It gives no indication of what approximations allow calculations of this power or efficiency, without an unphysical violation of fluctuation theorems and of the second law of thermodynamics. This work finds a microscopic symmetry which underlies the fluctuation theorems, beyond the Markovian quantum systems considered in Ref. [46]. This enables one to identify a family of approximations that allow tractable calculations of machine power and efficiency, with no risk of violating the second law or fluctuation theorems.

A. Overview of this work

The central observation of this work is the result connecting trajectories on the Keldysh contour in a system, to time-reversed trajectories in a time-reversed system, given in Sec. VII E and shown schematically in Fig. 5. The sections leading up to Sec. VII set the scene, with Sec. VI making the observation that such trajectories obey the first law of thermodynamics on average. Then Sec. VII itself provides a derivation of the result connecting trajectories to their time reverse.

The rest of this work then uses this result in deriving various fluctuation theorems. Section VIII uses it to derive various fluctuation theorems in various situations, such as the Jarzynski equality and the Crooks equation. In particular, it shows that the integral fluctuation theorem in Eq. (1) holds for any system which starts in a product state with the reservoirs, thereby showing that any such system obeys the second law of thermodynamics on average. Section IX provides similar proofs for situations where the system and reservoir start in a nonfactorizable initial state. Finally, Sec. X discusses approximations that respect the result in Sec. VII E, and thereby will not violate any of the fluctuation theorems in Secs. VIII and IX, and so will always satisfy the second law of thermodynamics on average.

B. A comment on the system-reservoir coupling

The recent literature on Keldysh for quantum thermodynamics [34–37] has strongly debated the role of the average energy stored in the system-reservoir coupling, $\langle E_{\text{s-r}}\rangle$, in models of adiabatic pumping or in which interactions are absent. The initial claim was that $\langle E_{\text{s-r}}\rangle$ should be separated into two equal parts, with one part being assigned to the system and the other part to the reservoir; however Ref. [36] argued that this was mainly a matter of calculational convenience.

In light of this debate, it is worth mentioning the role of $\langle E_{\text{s-r}}\rangle$ in the diagrammatic approach used here, whose differences from the approach in Refs. [34–37] are described in Sec. II below. First, $\langle E_{\text{s-r}}\rangle$ appears in the first law, but does not appear in the second law or the integral fluctuation theorem, since they involve entropy rather than energy. Second, it is *not* convenient to assign any part of $\langle E_{\text{s-r}}\rangle$ to the reservoirs for the following reason. In the cases considered here, each reservoir is in a state that is "simple"; that is to say it is in local equilibrium, which is completely described by two parameters: temperature and electrochemical potential. However, the system state is not "simple" in this sense, because it is typically far from equilibrium (due to the action of multiple reservoirs and/or driving), and requires more than just these two parameters to describe it. This means that the system-reservoir coupling is also not "simple." Hence, it is unhelpful to associate any part of the system-reservoir coupling with the reservoir state, because one then loses the simplicity of the latter. In contrast the energy in the system-reservoir coupling always appears together with the system energy (see Sec. VI), and as neither contribution is "simple" in the above sense, there is no disadvantage with making the choice to combine the two into a single *effective* internal system energy. That said, this article keeps the energy in the system-reservoir coupling separate from the system energy throughout, to avoid ambiguity.

II. HAMILTONIAN

This work considers a small quantum system with the Hamiltonian, $\hat{H}_{\text{sys}}(t)$, which may include a time-dependent driving and interactions between the particles in the system. This system [shown at the center of Fig. 1(a)] acts as a machine changing the heat and work in the reservoirs that surround it. Each term in $\hat{H}_{\text{sys}}(t)$ contains one creation operator for a system electronic state, $\hat{d}_i^\dagger$, for every annihilation operator, $\hat{d}_j$. This system is coupled to multiple reservoirs of noninteracting fermions (electrons) via couplings $\hat{V}_{\text{el}}^{(\alpha)}(t)$, or noninteracting bosons (photons or phonons) via couplings $\hat{V}_{\text{ph}}^{(\alpha)}(t)$. This article uses the word "setup" to refer to the system and reservoirs together; the total Hamiltonian of this setup is

$$\hat{H}_{\rm tot}(t) = \hat{H}_{\rm sys}(t) + \sum_{\alpha\in{\rm el}}\big[\hat{V}^{(\alpha)}_{\rm el}(t) + \hat{H}^{(\alpha)}_{\rm el}\big] + \sum_{\alpha\in{\rm ph}}\big[\hat{V}^{(\alpha)}_{\rm ph}(t) + \hat{H}^{(\alpha)}_{\rm ph}\big]. \tag{2}$$

The sums are over electron (el) and photon/phonon (ph) reservoirs. For el reservoirs,

$$\hat{H}^{(\alpha)}_{\rm el} = \sum_k E_{\alpha k}\hat{c}^\dagger_{\alpha k}\hat{c}_{\alpha k}, \tag{3}$$

for reservoir α's state k with energy, creation, and annihilation operators $E_{\alpha k}$, $\hat{c}^\dagger_{\alpha k}$, and $\hat{c}_{\alpha k}$. The tunnel coupling

$$\hat{V}^{(\alpha)}_{\rm el}(t) = \sum_k [\hat{V}^+_{\alpha k}(t)\hat{c}_{\alpha k} + \hat{V}^-_{\alpha k}(t)\hat{c}^\dagger_{\alpha k}], \tag{4}$$

where $\hat{V}^-_{\alpha k}(t)$ and $\hat{V}^+_{\alpha k}(t)$ contain only system operators, and may be time-dependent. The change in the system state when an electron is added from reservoir α's state k is given by $\hat{V}^+_{\alpha k}$. The reverse process is given by $\hat{V}^-_{\alpha k} = [\hat{V}^+_{\alpha k}]^\dagger$. The simplest case has $\hat{V}^+_{\alpha k} = \sum_i A^{(\alpha)}_{ik}\hat{d}^\dagger_i$; however if the coupling depends on the system state, then $\hat{V}^+_{\alpha k}$ contains extra factors of $\hat{d}^\dagger_j\hat{d}_{j'}$. For bosonic reservoirs, one replaces the fermionic operators $\hat{c}^\dagger_{\alpha k}$ and $\hat{c}_{\alpha k}$ with bosonic ones. The simplest case has $\hat{V}^+_{\alpha k} = \sum_{ij} A^{(\alpha)}_{ijk}\hat{d}^\dagger_i\hat{d}_j$, meaning the system goes from j to i when a boson is absorbed from reservoir α's state k.

The first step to using the real-time transport theory [27–30] is to write all system operators as $N \times N$ matrices acting on the basis of N many-body system states; see, e.g., Appendix C of Ref. [11]. We go to an interaction representation (indicated by calligraphic symbols), where system operators evolve under a matrix,

$$\mathcal{U}_{\rm sys}(\tau, t_0) = T \exp\left[-i\int_{t_0}^{\tau} H_{\rm sys}(t)dt\right], \tag{5}$$

with T indicating time ordering. Hence,

$$\mathcal{V}^\pm_{\alpha k}(\tau) = \mathcal{U}^\dagger_{\rm sys}(\tau;t_0)V^\pm_{\alpha k}(\tau)\mathcal{U}_{\rm sys}(\tau;t_0). \tag{6}$$

Reservoir operators evolve under $H^{(\alpha)}_{\rm el/ph}$, so we have

$$\begin{aligned}\hat{c}^\dagger_{\alpha k}(\tau) &= e^{iE_k(\tau-t_0)}\hat{c}^\dagger_{\alpha k},\\ \hat{c}_{\alpha k}(\tau) &= e^{-iE_k(\tau-t_0)}\hat{c}_{\alpha k}.\end{aligned} \tag{7}$$

The initial condition (at time t_0) is an arbitrary system state in a product state with the reservoirs. Each reservoir α is in its local equilibrium with temperature T_α and chemical potential μ_α ($\mu_\alpha = 0$ for reservoirs of photons or phonons). We treat $H_{\rm sys}$ exactly, and keep the reservoir's effect on the system finite, in the limit of vanishing reservoir level spacing. This requires taking the system's coupling to each reservoir mode to zero, as the density of such modes goes to infinity, so this coupling can be treated at lowest order (second order) [30,47–49]. Nonetheless, the system may interact with any number of reservoir modes at one time (all orders of cotunneling events), and these interactions do not commute. Upon tracing out the reservoirs, the resulting system dynamics are highly non-Markovian. Thus its dynamics are not described by Markovian master equations (Lindblad equations), whose thermodynamics have already been well studied [31]. These dynamics are represented in terms of a Keldysh double trajectory, as in Fig. 1(b), where each second-order interaction with a given reservoir mode is represented by a pair of interactions joined by a dashed line.

For readers familiar with the Keldysh methods reviewed in Kamenev's textbook [38] and used in Refs. [34–37], we note that the method used here is different at the level of what is treated as a perturbation. In Kamenev's textbook, the Hamiltonian is written in the single-particle basis; in this basis the Hamiltonian is quadratic in the absence of interactions between particles, and so is exactly soluble. One then uses increasingly sophisticated perturbative techniques to include the interaction terms such as electron-electron interactions (which are quartic in the single-particle operators). In contrast, this work uses a diagrammatic method on the Keldysh contour referred to as the *real-time transport theory* [27–30], which takes a different starting point: it starts in the many-body basis for the system Hamiltonian (the fact that it is in real time is not particularly important). In this basis, the physics of the system alone is trivial (including all interaction effects); however the system-reservoir couplings take forms that are too complicated to treat exactly. Thus, one has moved the difficulty from the interaction terms to the system-reservoir coupling terms. This is why this coupling must be treated as a perturbation, for which one sums up classes of irreducible diagrams within some suitable approximation scheme.

III. ASSUMPTION OF NO MAXWELL DEMONS IN RESERVOIRS

The equations of classical and quantum physics are reversible. For example, if all degrees of freedom in a quantum system were easy to observe and extract work from, the fact that the full wave function of the system and reservoirs undergoes unitary evolution means no (von Neumann) entropy is ever produced. In both classical and quantum physics, entropy production emerges from a physically motivated assumption about which degrees of freedom are easy to observe and extract work from, and which are not. Typically this assumption separates everything into macroscopic and microscopic dynamics, where macroscopic dynamics are easy to observe and extract work from, while the microscopic dynamics are inaccessible. All works on thermodynamics make some sort of assumption of this type, explicitly or implicitly.

The assumption at the basis of this work is presented here; for compactness it is referred to as the "assumption of no Maxwell demons in reservoirs." It requires that the system operate without knowing microscopic details of the reservoirs, beyond those encoded in the system-reservoir interaction in Eq. (2). For example, this disallows Maxwell's "observant and neat-fingered" demons [50] (which are usually just circuitry built by physicists) which measure individual reservoir states, and then feed back this information by making a change in the time dependence of $\hat{H}_{\rm sys}(t)$ or $\hat{V}^{(\alpha)}(t)$ in Eq. (2) which is conditional on the result of the measurement. Just as in classical mechanics, assuming no Maxwell

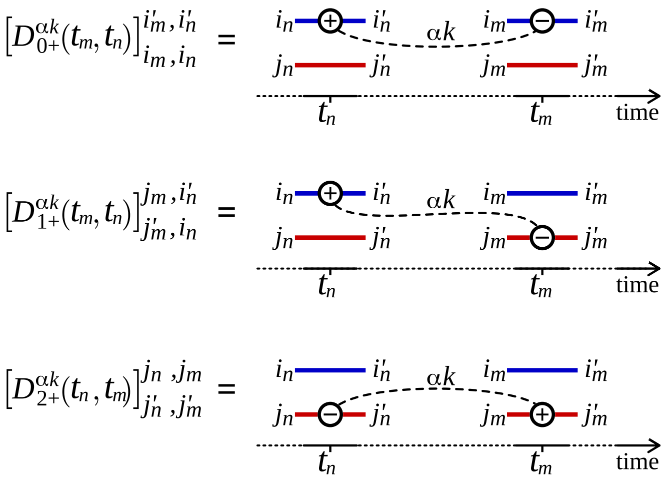
$[D^{\alpha k}_{0+}(t_m,t_n)]^{i'_m,i'_n}_{i_m,i_n}$ =
i_n i'_n αk i_m i'_m
j_n j'_n j_m j'_m
t_n t_m time
$[D^{\alpha k}_{1+}(t_m,t_n)]^{j_m,i'_n}_{j'_m,i_n}$ =
i_n i'_n αk i_m i'_m
j_n j'_n j_m j'_m
t_n t_m time
$[D^{\alpha k}_{2+}(t_n,t_m)]^{j_n,j_m}_{j'_n,j'_m}$ =
i_n i'_n αk i_m i'_m
j_n j'_n j_m j'_m
t_n t_m time

FIG. 2. The second-order interaction with reservoir α's mode k. Vertices marked ⊕ or ⊖ correspond to the matrices $\mathcal{V}^+_{\alpha k}$ or $\mathcal{V}^-_{\alpha k}$, respectively. The upper line is read from left to right, so the ⊕ vertex in D_{0+} or D_{1+} indicates the matrix element $[\mathcal{V}^+_{\alpha k}]_{i_{n+1} i_n}$. The lower line is read from right to left, so the ⊕ vertex in D_{2+} indicates $[\mathcal{V}^+_{\alpha k}]_{j_m j_{m+1}}$. Interaction D_{a-} is given by D_{a+} with ⊕ ↔ ⊖, for $a = 0, 1, 2$.

demons is crucial in the emergence of the second law from the underlying theory. This assumption makes all classical correlations and quantum entanglement between system and reservoirs at the end of the evolution irrelevant, since the system cannot extract work from them. Hence, one can trace out the reservoirs when calculating system entropy, and vice versa. Further, even though the system pushes certain reservoir modes out of equilibrium, it is assumed that this information is inaccessible, so no more work can be extracted from the reservoir than if it were in a thermal state with the same energy.

Superficially, one might think this work has nothing to say about experimental implementations of Maxwell demons in quantum systems, similarly to Refs. [51,52]. However, in those cases where the demon is completely mechanical (made of some finite number of degrees of freedom coupled to reservoirs with or without time-dependent driving), we can include these degrees of freedom in the system Hamiltonian $\hat{H}_{\rm sys}$, and all results in this work apply.

IV. TRAJECTORIES

Consider a trajectory γ on the Keldysh contour, whose upper line goes from the system's many-body state i_0 at time t_0 to i at time t, and whose lower line goes from j_0 to j [see examples in Fig. 5(a)]. Matrix elements for transitions are time ordered on the upper line and reverse-time ordered on the lower line. Each transition (each dashed line in γ) has a weight determined by whether it is $D_{0\pm}$, $D_{1\pm}$, or $D_{2\pm}$ in Fig. 2 (see below). Real transitions correspond to D_{1+} in Fig. 2 and virtual transitions to D_{0+} and D_{2+}. The trajectory's weight, $P(\gamma)$, is the product of all of these factors of $D_{a\pm}$, multiplied by a factor of -1 for each crossing of dashed lines [30]. The probability to go from one system state to another in time t is simply the sum of the weights of all trajectories between those states.

The dashed lines have the following weights,

$$\big[D^{\alpha k}_{0+}\big]^{i'_m,i'_n}_{i_m,i_n} = -[\mathcal{V}^-_{\alpha k}(t_m)]^{i'_m}_{i_m}[\mathcal{V}^+_{\alpha k}(t_n)]^{i'_n}_{i_n} f^+_{\alpha k} e^{i\Phi^{mn}_k}, \tag{8}$$

$$\big[D^{\alpha k}_{1+}\big]^{j_m,i'_n}_{j'_m,i_n} = [\mathcal{V}^-_{\alpha k}(t_m)]^{j_m}_{j'_m}[\mathcal{V}^+_{\alpha k}(t_n)]^{i'_n}_{i_n} f^+_{\alpha k} e^{i\Phi^{mn}_k}, \tag{9}$$

$$\big[D^{\alpha k}_{2+}\big]^{j_n,j_m}_{j'_n,j'_m} = -[\mathcal{V}^-_{\alpha k}(t_n)]^{j_n}_{j'_n}[\mathcal{V}^+_{\alpha k}(t_m)]^{j_m}_{j'_m} f^+_{\alpha k} e^{i\Phi^{mn}_k}, \tag{10}$$

where $[\mathcal{V}]^{i'}_i = \langle i'|\mathcal{V}|i\rangle$ and $\Phi^{mn}_k = E_k(t_m - t_n)$. The factor $f^+_{\alpha k}$ is the number of particles in state k of reservoir α; it is $f^+_{\alpha k} = 1/(e^{\delta S_{\alpha k}} + \nu)$ with $\nu = 1$ for fermionic reservoirs and $\nu = -1$ for bosonic reservoirs. Here [18],

$$\delta S_{\alpha k} = (E_k - \mu_\alpha)/T_\alpha, \tag{11}$$

which is the entropy change of reservoir α when a particle is added to state k. Identifying Eq. (11) with an entropy change follows from the Claussius definition of entropy, applicable here because each reservoir is in its own local thermodynamics equilibrium with a well-defined temperature.

The weight of D_{a-} (for $a = 0, 1, 2$) is given by the Hermitian conjugate of D_{a+} (so $\mathcal{V}^+ \leftrightarrow \mathcal{V}^-$ and $i\Phi^{mn}_k \to -i\Phi^{mn}_k$) with $f^+_{\alpha k}$ replaced by $f^-_{\alpha k}$. Hence

$$\big[D^{\alpha k}_{0-}\big]^{i'_m,i'_n}_{i_m,i_n} = -[\mathcal{V}^+_{\alpha k}(t_m)]^{i'_m}_{i_m}[\mathcal{V}^-_{\alpha k}(t_n)]^{i'_n}_{i_n} f^-_{\alpha k} e^{i\Phi^{nm}_k}, \tag{12}$$

$$\big[D^{\alpha k}_{1-}\big]^{j_m,i'_n}_{j'_m,i_n} = [\mathcal{V}^+_{\alpha k}(t_m)]^{j_m}_{j'_m}[\mathcal{V}^-_{\alpha k}(t_n)]^{i'_n}_{i_n} f^-_{\alpha k} e^{i\Phi^{nm}_k}, \tag{13}$$

$$\big[D^{\alpha k}_{2-}\big]^{j_n,j_m}_{j'_n,j'_m} = -[\mathcal{V}^+_{\alpha k}(t_n)]^{j_n}_{j'_n}[\mathcal{V}^-_{\alpha k}(t_m)]^{j_m}_{j'_m} f^-_{\alpha k} e^{i\Phi^{nm}_k}. \tag{14}$$

Here $f^-_{\alpha k}$ is the number of ways one can add a particle to state k of reservoir α. For any reservoir (fermionic, bosonic, or other) in internal equilibrium,

$$f^-_{\alpha k} = e^{\delta S_{\alpha k}} f^+_{\alpha k}, \tag{15}$$

which is known as local detailed balance or microreversibility. For fermion or boson distributions, this is guaranteed by the fact that $f^-_{\alpha k} = 1 + \nu f^+_{\alpha k}$ with $\nu = +1$ for fermions, and $\nu = -1$ for bosons. Physically, $D^{\alpha k}_{1-}$ removes a particle from the system and adds it to state k of reservoir α; this adds a work of μ_α and heat of $(E_k - \mu_\alpha)$ to reservoir α. Thus, $D^{\alpha k}_{1-}$ involves a change of reservoir α's entropy of $\delta S_{\alpha k}$ in Eq. (11). The reverse process, $D^{\alpha k}_{1+}$, removes such a particle from reservoir α, changing the reservoir's entropy by $-\delta S_{\alpha k}$. Contributions $D^{\alpha k}_{0\pm}$ and $D^{\alpha k}_{2\pm}$ do not change the number of particles in the reservoirs, and so involve no reservoir entropy change.

V. TOTAL ENTROPY

The assumption of no Maxwell demons in the reservoirs implies that entanglement between system and reservoir cannot be used to produce work. Then the correct definition of the total entropy production, $\Delta S_{\rm tot}$, is the sum of that for the system (sys) and reservoirs (res),

$$\Delta S_{\rm tot} = \Delta S_{\rm sys} + \sum_\alpha \Delta S^{(\alpha)}_{\rm res}, \tag{16}$$

with no term related to system-reservoir entanglement. We take the change in entropy of each reservoir to be given by the Claussius formula. This means that the change in reservoir α's entropy, $\Delta S^{(\alpha)}_{\rm res}$, for a trajectory γ is taken to be the sum of the entropy changes $\mp\delta S_{\alpha k}$ associated with each of the $D_{1\pm}^{\alpha k}$ transitions in γ.

As the system is typically in a highly nonequilibrium state, one cannot use the Claussius law to calculate its entropy. In the stochastic thermodynamics of classical systems [8,10,53], an entropy is assigned to each system state in such a way that the entropy of the system averaged over all such system states is the Shannon entropy. For quantum systems, one can do exactly the same thing, if (and only if) the system's density matrix is in its diagonal basis. To get this entropy for the system's initial density matrix (at time t_0), we write it as

$$\rho_{ml}^{\rm sys}(t_0) = \sum_{n_0} [\mathcal{W}_0]_{mn_0} P_{n_0}(t_0) [\mathcal{W}_0^\dagger]_{n_0 l}, \tag{17}$$

where $\mathcal{W}_0$ is the unitary matrix which rotates the system density matrix at time t_0 to its diagonal basis. This means that $P_{n_0}(t_0)$ is the probability to find the system in state n_0 of its diagonal basis. In this basis, the system's von Neumann entropy, $-{\rm tr}\{\rho_{\rm sys}(t_0)\ln[\rho_{\rm sys}(t_0)]\}$, is simply $-\sum_{n_0} P_{n_0}(t_0)\ln[P_{n_0}(t_0)]$, where the sum is over the elements of the diagonal density matrix. Thus, one can treat each element in the sum as a contribution to the entropy from a given initial state, so that state n_0's contribution to the entropy of the initial system state is

$$S_{n_0}(t_0) = -\ln[P_{n_0}(t_0)], \tag{18}$$

with the average over all n_0 (i.e., a sum over n_0 weighted by the probability of state n_0) giving the system's initial von Neumann entropy. The final system state's entropy (at time t) is calculated in the same way by rotating to the diagonal basis of the final system density matrix, given by $\rho_{ml}^{\rm sys}(t) = \sum_n \mathcal{W}_{mn} p_n(t) [\mathcal{W}^\dagger]_{nl}$ for unitary $\mathcal{W}$, and assigning to the state n an entropy of

$$S_n(t) = -\ln[P_n(t)], \tag{19}$$

with $P_n(t)$ being the probability that the system is in state n of the diagonal basis of its reduced density matrix at time t. Equations (18), (19) can be used to associate trajectory $\gamma_{\rm d}$, from initial state n_0 to final state n, with an entropy change in the system of

$$\Delta S_{\rm sys}(\gamma) = S_n(t) - S_{n_0}(t_0) = -\ln\left[\frac{P_n(t)}{P_{n_0}(t_0)}\right], \tag{20}$$

as in the stochastic thermodynamics of classical rate equations. Recall that this is only possible because the trajectory $\gamma_{\rm d}$ is defined as going from a system state n_0 in the diagonal basis of the system density matrix at time t_0 to a system state n in the diagonal basis of the system's final density matrix (which is found by tracing out the reservoirs at the end of the evolution). This requires calculating the final density matrix (and finding its diagonal basis); this is much like in the usual stochastic thermodynamics, where one also needs a complete knowledge of the final-state probability distribution to assign entropies to it.

VI. FIRST LAW OF THERMODYNAMICS

Here we show that energy conservation ensures that the first law of thermodynamics is obeyed on average. If one goes beyond the average, there are fluctuations that violate the first law, much like the fluctuations that violate the second law. These are little studied to date and merit a detailed study of their own. In this section, we restrict ourselves to considering the average energy in the setup, and thereby show that the first law holds on average.

To ensure energy conservation, one must sum the three terms which contribute to the total energy: the energy in the reservoirs, the energy in the system, and the energy in the system-reservoir coupling. If the system is not driven this total energy is conserved. If the system is driven then the difference between the final and initial total energy is the work done by the drive; thus between time t_0 and time t, the average work done by the drive is

$$\begin{aligned}\langle\Delta W_{\rm drive}(t;t_0)\rangle = \langle\Delta E_{\rm res}(t;t_0)\rangle + \langle\Delta E_{\rm sys}(t;t_0)\rangle \\ + \langle\Delta E_{\rm s\text{-}r}(t;t_0)\rangle,\end{aligned} \tag{21}$$

where $\Delta E_{\rm res}(t;t_0)$ is the energy change in the reservoirs, $\Delta E_{\rm sys}(t;t_0) = E_{\rm sys}(t) - E_{\rm sys}(t_0)$ is the energy change in the system, and $\Delta E_{\rm s\text{-}r}(t;t_0) = E_{\rm s\text{-}r}(t) - E_{\rm s\text{-}r}(t_0)$ is the energy change in the system-reservoir coupling.

Of course, all terms in this sum are necessary to get to get energy conservation, irrespective of whether one can physically measure each of them or not. Under the assumption of no Maxwell demons in the reservoirs, made in Sec. III, one can measure the energy in the system and reservoirs, but not that in the system-environment coupling, since that depends on the state of individual reservoir modes. In such a case, one could none the less determine $\langle\Delta E_{\rm s\text{-}r}(t;t_0)\rangle$ by using Eq. (21), assuming one can also measure the work done by the drive, $\langle\Delta W_{\rm drive}(t;t_0)\rangle$.

The average energy in the quantum system is

$$\langle E_{\rm sys}(t)\rangle = {\rm tr}_{\rm sys}[\hat{H}_{\rm sys}(t)\rho_{\rm sys}(t)], \tag{22}$$

while that in the system-reservoir coupling is

$$\begin{aligned}\langle E_{\rm s\text{-}r}(t)\rangle = \sum_{\alpha\in{\rm el}} {\rm tr}\big[\hat{V}_{\rm el}^{(\alpha)}(t)\hat{\rho}_{\rm tot}(t)\big] \\ + \sum_{\alpha\in{\rm ph}} {\rm tr}\big[\hat{V}_{\rm ph}^{(\alpha)}(t)\hat{\rho}_{\rm tot}(t)\big],\end{aligned} \tag{23}$$

where the trace in $\langle E_{\rm sys}\rangle$ is over the system states and $\rho_{\rm sys}(t)$ is the reduced system density matrix, but $\langle E_{\rm s\text{-}r}\rangle$ contains traces over the total density matrix (including reservoirs).

The trajectories which sum to give the average change in the system energy, $\langle\Delta E_{\rm sys}(t;t_0)\rangle$, are those considered elsewhere in this article, such as those in Fig. 1(b) or Fig. 5. However, the trajectories which sum to give the average change in the energy in the system-reservoir coupling, $\langle\Delta E_{\rm s\text{-}r}(t;t_0)\rangle$, are rather different from those considered elsewhere in this article. They have an additional single interaction vertex with mode k of reservoir α at some time before t, so that at time t the system is in a superposition of a state with different numbers of particles in reservoir α (see Fig. 3). Luckily, they have exactly the same structure as those used to calculate the current into the system in Refs. [27–30]; currents are given the difference between the term that creates a particle in the reservoir and one that destroys a particle, while the energy in the system-reservoir coupling is given by the sum of these two terms. These trajectories are

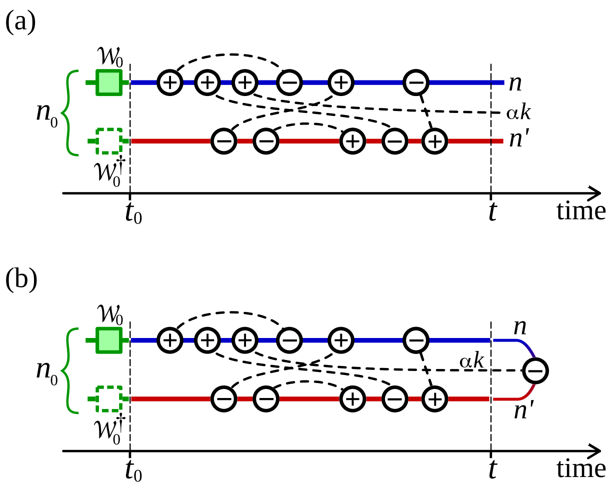

FIG. 3. (a) A trajectory which contributes to $[\mathcal{V}^{-}_{\alpha k}(t)]^{n'}_{n}\hat{c}^{\dagger}_{\alpha k}$, and (b) its contribution to the *average* energy in the system-reservoir coupling, $\langle E_{\text{s-r}}\rangle$. Such contributions to $\langle E_{\text{s-r}}\rangle$ are the same as those necessary to calculate the current at time t, and their evaluation has been greatly discussed in Refs. [27–30]. (a) A trajectory giving a non-zero sys-res coupling energy (b) its contribution to $\langle E_{\text{s-r}}\rangle$

not discussed further here, because Refs. [27–30] go into great detail about how to calculate their contribution.

The fact that reservoir α is in local thermodynamic equilibrium defined by its temperature T_α and electrochemical potential μ_α means it makes sense to split its energy, $\langle E^{(\alpha)}_{\text{res}}(t)\rangle = \text{tr}[H^{(\alpha)}_{\text{el/ph}}\hat{\rho}(t)]$, into two contributions

$$\big\langle E^{(\alpha)}_{\text{res}}(t)\big\rangle = -\big\langle W^{(\alpha)}_{\text{res}}(t)\big\rangle - \big\langle Q^{(\alpha)}_{\text{res}}(t)\big\rangle, \tag{24}$$

with

$$\big\langle W^{(\alpha)}_{\text{res}}(t)\big\rangle = -\text{tr}\left[\mu_\alpha \sum_k \hat{c}^{\dagger}_{\alpha k}\hat{c}_{\alpha k}\hat{\rho}(t)\right], \tag{25}$$

$$\big\langle Q^{(\alpha)}_{\text{res}}(t)\big\rangle = -\text{tr}\left[\sum_k (E_{\alpha k}-\mu_\alpha)\hat{c}^{\dagger}_{\alpha k}\hat{c}_{\alpha k}\hat{\rho}(t)\right]. \tag{26}$$

Then the change in reservoir energy between time t_0 and t can be defined as

$$\big\langle \Delta E^{(\alpha)}_{\text{res}}(t;t_0)\big\rangle = -\big\langle \Delta W^{(\alpha)}_{\text{res}}(t;t_0)\big\rangle - \big\langle \Delta Q^{(\alpha)}_{\text{res}}(t;t_0)\big\rangle; \tag{27}$$

the first quantity here is the average work done by the reservoir $\langle \Delta W^{(\alpha)}_{\text{res}}(t;t_0)\rangle = \langle W^{(\alpha)}_{\text{res}}(t)\rangle - \langle W^{(\alpha)}_{\text{res}}(t_0)\rangle$, and the second quantity is the average heat flow out of the reservoir $\langle \Delta Q^{(\alpha)}_{\text{res}}(t;t_0)\rangle = \langle Q^{(\alpha)}_{\text{res}}(t)\rangle - \langle Q^{(\alpha)}_{\text{res}}(t_0)\rangle$. There is no ambiguity in this separation, because the fact that the reservoir is in local equilibrium means that the former (the work done) has no entropy change associated with it, while the latter (the heat change) is associated with an entropy change of

$$\Delta S^{(\alpha)}_{\text{res}} = \Delta Q^{(\alpha)}_{\text{res}}\big/T_\alpha. \tag{28}$$

The trajectories which sum to give these average changes in a reservoir's energy are those considered elsewhere in this article, such as those shown in Fig. 1(b) or Fig. 5. However the change in work or heat in the reservoir α is extremely easy to read from a given trajectory; one simply sums up the change in work or heat for each dashed line symbolizing $D^{\alpha k}_{1\pm}$, as outlined at the end of Sec. IV.

Given these definitions and Eq. (21), one easily arrives at the first law of thermodynamics for the average dynamics of the setup:

$$\begin{aligned}&\langle \Delta E_{\text{sys}}(t;t_0)\rangle + \langle \Delta E_{\text{s-r}}(t;t_0)\rangle\\ &\quad= \langle \Delta W(t;t_0)\rangle + \sum_\alpha \big\langle \Delta Q^{(\alpha)}_{\text{res}}(t;t_0)\big\rangle,\end{aligned} \tag{29}$$

where we define $\Delta W(t;t_0)$ as the total work done on the system by drive or reservoirs,

$$\langle \Delta W(t;t_0)\rangle = \langle \Delta W_{\text{drive}}(t;t_0)\rangle + \sum_\alpha \big\langle \Delta W^{(\alpha)}_{\text{res}}(t;t_0)\big\rangle. \tag{30}$$

It thereby seems natural to interpret $\langle \Delta E_{\text{sys}}(t;t_0)\rangle + \langle \Delta E_{\text{s-r}}(t;t_0)\rangle$ as the change in the *effective* internal energy of the system (an effective energy which includes the system-reservoir coupling), as mentioned in Sec. I B.

Just as in classical thermodynamics systems, the simplest cases to consider are those where the system returns to its initial state at the end of the evolution, so that its internal energy is the same at the final time, t, as it was at the initial time, t_0. Then $\langle \Delta E_{\text{sys}}(t;t_0)\rangle = \langle \Delta E_{\text{s-r}}(t;t_0)\rangle = 0$, which means that Eq. (29) directly gives the simplest and best-known consequence of the first law: the work output of the machine equals the heat absorbed from the reservoirs.

Note that the change of energy in the reservoirs was separated into a change of heat and a change of work, but this was not done for the energy of the system or the system-reservoir coupling. The reason is that each reservoir is in local thermodynamic equilibrium, with a well-defined temperature, when the system and system-reservoir couplings are typically far from equilibrium with no well-defined temperature. Thus there is no ambiguity in the separation of energy into heat and work in a reservoir, see Eq. (28), but there is no simple way to make the same separation for the system or for the system-reservoir couplings.

VII. TIME-REVERSED SETUP

As in classical systems, one derives fluctuation theorems by comparing two different setups (A and B), where the Hamiltonian in setup B is the time reverse of the Hamiltonian in setup A over the time window from t_0 to t. To be clear, whatever the Hamiltonian of setup A, we can invent a setup B whose Hamiltonian is the time reverse of setup A. In the special case of a time-independent Hamiltonian without external magnetic fields or spins, the two setups are identical, but otherwise they are not.

The objective of this section is to make the connection between weight of trajectories on the Keldysh contour in setup B and setup A. This starts by making the connection between the terms in the Hamiltonians of setups A and B in Sec. VII A, and then between the perturbative terms in the interaction representation in Sec. VII C. This enables one to make the connection between the weight of individual transitions in Sec. VII D, from which one gets the connection between weight of trajectories on the Keldysh contour in setup B and setup A in Sec. VII E. The *central observation* of this work is this

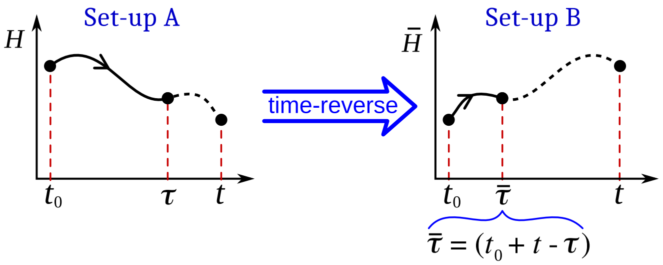

FIG. 4. A sketch of how time reversal affects the Hamiltonian, in the absence of external magnetic fields and spins. If there are external magnetic fields and spins, then the time reverse is given in the Appendix.

relationship, given in Eq. (48) below. It is this relationship which is so similar to the relationship between trajectories in the stochastic thermodynamics theory for classical Markovian systems [8,10,53] that we can use very similar logic to derive various well-known fluctuation theorems in Sec. VIII.

A. Time-reversed Hamiltonian

If A's Hamiltonian (system+reservoir) is $\hat{H}$ in Eq. (2), then B's is $\overline{\hat{H}(\tau)} = \hat{\Theta}^\dagger \hat{H}(t_0 + t - \tau)\hat{\Theta}$, where $\hat{\Theta}$ is the time-reverse operator in Messiah's texbook [54]. The main results we need from Messiah's texbook are recalled here in the Appendix, with the most trivial case sketched in Fig. 4. Thus if setup A has a given time-dependent system Hamiltonian, $H_{\rm sys}(\tau)$, with given system-reservoir couplings, $V^\pm_{\alpha k}(\tau)$, then setup B is *chosen* to have the system Hamiltonian and system-reservoir couplings

$$\overline{\hat{H}_{\rm sys}(t_0 + t - \tau)} = \hat{\Theta}^\dagger \hat{H}_{\rm sys}(\tau)\hat{\Theta}, \tag{31}$$

$$\overline{\hat{V}^\pm_{\alpha k}(t_0 + t - \tau)} = \hat{\Theta}^\dagger \hat{V}^\pm_{\alpha k}(\tau)\hat{\Theta}, \tag{32}$$

where the bar above a symbol means that it is in setup B, while the bar's absence means it is in setup A.

These equations are cast in terms of matrix elements by inserting them between $\langle \bar{\imath}| = \langle i|\hat{\Theta}$ and $|\bar{\jmath}\rangle = \hat{\Theta}^\dagger|j\rangle$; then

$$[\,\overline{H_{\rm sys}(t_0 + t - \tau)}]^{\bar{\imath}}_{\bar{\jmath}} = [H_{\rm sys}(\tau)]^i_j, \tag{33}$$

$$[\,\overline{V^\pm_{\alpha k}(t_0 + t - \tau)}]^{\bar{\imath}}_{\bar{\jmath}} = [V^\pm_{\alpha k}(\tau)]^i_j, \tag{34}$$

where $[\cdots]^{\bar{\imath}}_{\bar{\jmath}} = \langle \bar{\imath}|\cdots|\bar{\jmath}\rangle$. Thus the matrix elements for transitions from system state $|\bar{\jmath}\rangle$ to system state $|\bar{\imath}\rangle$ in setup B (whose Hamiltonian is the time reverse of setup A's) are the same as the matrix elements from system state $|j\rangle$ to system state $|i\rangle$ in setup A. Spinless systems written in a basis of position states are trivial, because then $|\bar{\imath}\rangle = |i\rangle$. However, if one is working with basis states with nonzero momentum states, then $|\bar{\imath}\rangle$ is the state with the opposite momentum from $|i\rangle$. If one is working with spins, then the state $|\bar{\imath}\rangle$ is the state with the opposite spin from from state $|i\rangle$.

Equally, the reservoir Hamiltonians for setup B are the time reverse of those in setup A, so reservoir α in setup B has a Hamiltonian

$$\overline{\hat{H}^{(\alpha)}_{\rm el}} = \hat{\Theta}^\dagger \hat{H}^{(\alpha)}_{\rm el}\hat{\Theta}, \tag{35}$$

where $\hat{H}^{(\alpha)}_{\rm el}$ is that reservoir's Hamiltonian in setup A. In the absence of spins or external magnetic fields, this time-reverse operation is of no consequence. However, if reservoir α in setup A is a reservoir of electrons which are spin-up with respect to some axis, then the same reservoir in setup B will contain electrons which are spin-down with respect to that axis. Similarly, if there is an external magnetic field acting on the reservoir in setup A, then the field must be reversed in that reservoir in setup B. For photon or phonon reservoirs, the relation between their Hamiltonians in the two setups is the same as in Eq. (35).

B. Reservoir states are not time reversed

If one evolves an initial state under a Hamiltonian, time-reverses the state, and evolves it under the time-reversed Hamiltonian, the dynamics in the second part of the evolution will look like a time reverse of the dynamics in the first part of the evolution.

However, this work's setup A and setup B are a different time-reversal situation, in which each setup is divided into a system and reservoirs, and we make the "assumption of no Maxwell demons in the reservoirs" in Sec. III. This assumes the setup and its drive are not aware of the microscopic dynamics of the reservoirs; as such one cannot time-reverse the state of the individual modes in the reservoirs, even if one can can time-reverse the reservoirs' Hamiltonians (typically time-reversing the reservoir part of the Hamiltonian only requires interchanging the chemical potentials on spin-up and spin-down reservoirs and reversing any external magnetic fields acting on the reservoirs). Thus, even if we time-reverse the system state and time-reverse the total Hamiltonian, we will not see time-reversed dynamics, because we have not time-reversed the reservoir states.

Suppose setup A starts with the system and reservoirs in a product state, and then evolves. The system becomes correlated and/or entangled with individual reservoir modes. A measurement of the system state indicates that it is decohering and decaying towards a thermal state. A measurement of individual reservoir modes shows that an infinitesimal proportion of them are acquiring a nonthermal state. Then in setup B, the total Hamiltonian is the time reverse of that in setup A, and the initial state is a product state, where the system state is the time reverse of the final system state in setup A, and the reservoir modes are taken to be thermal (i.e., not time reversed). The system state in setup B does *not* become less correlated and/or entangled with the reservoirs as it evolves (as it would if we had time-reversed the full state, including the reservoir modes). Instead, the system continues to become more entangled with reservoir modes, which means that a measurement of the system state in setup B will indicate that it also decoheres and decays towards a thermal state.

C. Time reversal for the interaction representation

Under time reversal, the matrix representation of the system evolution operator is

$$\overline{\mathcal{U}_{\rm sys}(t + t_0 - \tau; t_0)} = \hat{\Theta}^\dagger \mathcal{U}^\dagger_{\rm sys}(t;\tau)\hat{\Theta}. \tag{36}$$

Now, to simplify the algebra, it is assumed that a complete solution of the dynamics under $H_{\rm sys}$ exists; then the final state of the system (its state at given time t) can always be written in a basis chosen such that $\mathcal{U}_{\rm sys}(t;t_0)=1$. Then, the unitary of $\mathcal{U}_{\rm sys}$ means that

$$\mathcal{U}^\dagger_{\rm sys}(t;\tau)\ =\ \mathcal{U}_{\rm sys}(\tau;t_0). \tag{37}$$

Given this one has

$$\overline{\mathcal{U}_{\rm sys}(t+t_0-\tau;t_0)} = \hat\Theta^\dagger\,\mathcal{U}_{\rm sys}(\tau;t_0)\,\hat\Theta. \tag{38}$$

The interaction between the system and the reservoirs at time $\bar\tau=(t_0+t-\tau)$ is written in the interaction representation for the *time-reversed Hamiltonian* as

$$\overline{\mathcal{V}^\pm_{\alpha k}(\bar\tau)} = \overline{\mathcal{U}^\dagger_{\rm sys}(\bar\tau;t_0)}\ \overline{V^\pm_{\alpha k}(\bar\tau)}\ \overline{\mathcal{U}_{\rm sys}(\bar\tau;t_0)}, \tag{39}$$

where the matrix $\overline{V^\pm_{\alpha k}(\bar\tau)}$ is defined above. Substituting in Eqs. (32), (38) on the right, and comparing with Eq. (6), one finds that

$$\overline{\mathcal{V}^\pm_{\alpha k}(t_0+t-\tau)} = \hat\Theta^\dagger\mathcal{V}^\pm_{\alpha k}(\tau)\hat\Theta \tag{40}$$

for all τ between t_0 and t.

For what follows it is convenient to cast this equality in terms of matrix elements by inserting it between $\langle\bar\imath| = \langle i|\hat\Theta$ and $|\bar\jmath\rangle = \hat\Theta^\dagger|j\rangle$; then

$$[\overline{\mathcal{V}^\pm_{\alpha k}(t_0+t-\tau)}]^{\bar\imath}_{\bar\jmath}\ =\ [\mathcal{V}^\pm_{\alpha k}(\tau)]^i_j. \tag{41}$$

Thus the matrix element for reservoir-induced transitions from system state $|\bar\jmath\rangle$ to system state $|\bar\imath\rangle$ in setup B (the setup whose Hamiltonian is the time reverse of setup A's) is the same as the matrix element from system state $|j\rangle$ to system state $|i\rangle$ in setup A.

D. Time-reversal symmetry between $D^{\alpha k}_{a\pm}$ transitions

Equation (9) implies that the D_{1+} transition in the time-reserved system (setup B) must have the weight

$$\big[\overline{D^{\alpha k}_{1+}(\bar t_m,\bar t_n)}\big]^{\bar\jmath_m,\bar\imath'_n}_{\bar\jmath'_m,\bar\imath_n} = [\overline{\mathcal{V}^-_{\alpha k}(\bar t_m)}]^{\bar\jmath_m}_{\bar\jmath'_m}[\overline{\mathcal{V}^+_{\alpha k}(\bar t_n)}]^{\bar\imath'_n}_{\bar\imath_n} \times f^+_{\alpha k}\exp[iE_k(\bar t_m-\bar t_n)], \tag{42}$$

where $\bar t_n = t_0+t-t_n$. Now substituting in Eq. (41) and noting that $(\bar t_m-\bar t_n) = (t_n-t_m)$, we get

$$\big[\overline{D^{\alpha k}_{1+}(\bar t_m,\bar t_n)}\big]^{\bar\jmath_m,\bar\imath'_n}_{\bar\jmath'_m,\bar\imath_n} = [\mathcal{V}^-_{\alpha k}(t_m)]^{j_m}_{j'_m}[\mathcal{V}^+_{\alpha k}(t_n)]^{i'_n}_{i_n} \times f^+_{\alpha k}\exp[iE_k(t_n-t_m)]. \tag{43}$$

Now comparing this with D_{1-} in Eq. (13), one sees the only difference is the factors of $f^\pm_{\alpha k}$. However, local detailed balance in reservoir α implies Eq. (15), so

$$\big[\overline{D^{\alpha k}_{1+}(\bar t_m,\bar t_n)}\big]^{\bar\jmath_m,\bar\imath'_n}_{\bar\jmath'_m,\bar\imath_n} = \big[D^{\alpha k}_{1-}(t_m,t_n)\big]^{j_m,i'_n}_{j'_m,i_n}e^{-\delta S_{\alpha k}}. \tag{44}$$

Exactly the same logic holds if one starts with $\overline{D^{\alpha k}_{1-}}$ in place of $\overline{D^{\alpha k}_{1+}}$. One just has to take the Hermitian conjugate throughout (so $\mathcal{V}^+\leftrightarrow\mathcal{V}^-$ and $i\Phi^{mn}_k\to -i\Phi^{mn}_k$) and replace $f^+_{\alpha k}$ by $f^-_{\alpha k}$, getting the results in Eq. (47a).

Similarly, Eq. (8) means that

$$\big[\overline{D^{\alpha k}_{0+}(\bar t_n,\bar t_m)}\big]^{\bar\jmath_n,\bar\jmath_m}_{\bar\jmath'_n,\bar\jmath'_m} = -[\overline{\mathcal{V}^-_{\alpha k}(\bar t_n)}]^{\bar\jmath_n}_{\bar\jmath'_n}[\overline{\mathcal{V}^+_{\alpha k}(\bar t_m)}]^{\bar\jmath_m}_{\bar\jmath'_m} \times f^+_{\alpha k}\exp[iE_k(\bar t_n-\bar t_m)]; \tag{45}$$

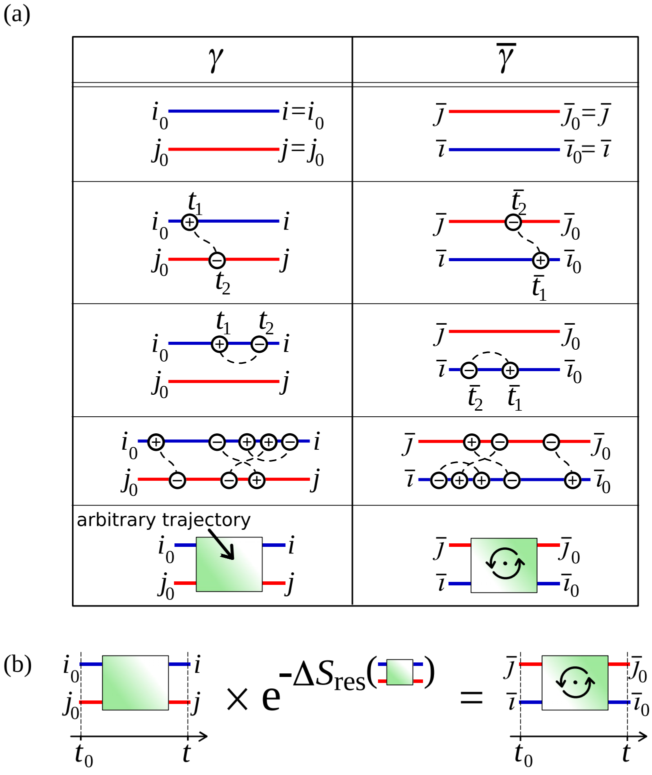
(a)
γ
$\overline{\gamma}$
i_0 $i=i_0$ j_0 $j=j_0$
$\bar\jmath$ $\bar\jmath_0=\bar\jmath$ $\bar\imath$ $\bar\imath_0=\bar\imath$
t_1 t_2 i_0 i j_0 j
$\bar t_2$ $\bar t_1$ $\bar\jmath$ $\bar\jmath_0$ $\bar\imath$ $\bar\imath_0$
arbitrary trajectory
(b)
i_0 j_0 i j t_0 t
$\times\ e^{-\Delta S_{\rm res}}$
=
$\bar\jmath$ $\bar\imath$ $\bar\jmath_0$ $\bar\imath_0$ t_0 t

FIG. 5. (a) Time reversal of trajectories on the Keldysh contour, via a 180° rotation in the plane of the page. The interaction times are related by $\bar t_n = t+t_0-t_n$. (b) A graphical representation of Eq. (48), where the shaded box is the trajectory's weight, $P(\gamma)$, and its 180° rotation is $\overline{P}(\overline{\gamma})$.

note that $\bar t_n > \bar t_m$, since it is assumed that $t_m > t_n$. As above, Eq. (41) is substituted in, and one notes that $(\bar t_m-\bar t_n) = (t_n-t_m)$, to get

$$\big[\overline{D^{\alpha k}_{0+}(\bar t_n,\bar t_m)}\big]^{\bar\jmath_n,\bar\jmath_m}_{\bar\jmath'_n,\bar\jmath'_m} = -[\mathcal{V}^-_{\alpha k}(t_n)]^{j_n}_{j'_n}[\mathcal{V}^+_{\alpha k}(t_m)]^{j_m}_{j'_m} \times f^+_{\alpha k}\exp[iE_k(t_m-t_n)], \tag{46}$$

which is the same as the right-hand side of Eq. (10). One can do the same for $\overline{D^{\alpha k}_{0-}}$.

The result of all these relations between D's in the time-reversed setup (setup B) and the original setup (setup A) can be summarized as follows:

$$\big[\overline{D^{\bar\alpha k}_{1\mp}(\bar t_m,\bar t_n)}\big]^{\bar\jmath_m,\bar\imath'_n}_{\bar\jmath'_m,\bar\imath_n} = \big[D^{\alpha k}_{1\pm}(t_n,t_m)\big]^{i'_n,j_m}_{i_n,j'_m}e^{\pm\delta S_{\alpha k}}, \tag{47a}$$

$$\big[\overline{D^{\bar\alpha k}_{0\pm}(\bar t_n,\bar t_m)}\big]^{\bar\jmath_n,\bar\jmath_m}_{\bar\jmath'_n,\bar\jmath'_m} = \big[D^{\alpha k}_{2\pm}(t_n,t_m)\big]^{j_n,j_m}_{j'_n,j'_m}, \tag{47b}$$

where $\bar t_n = t_0+t-t_n$.

E. Time-reversed trajectories

For any trajectory γ on the Keldysh contour in setup A, one can define a trajectory $\overline{\gamma}$ in setup B which is the time reverse of γ. More precisely, $\overline{\gamma}$ is defined by rotating γ by 180° in the plane of the page and replacing all states by their time reverse; see Fig. 5(a). The time reverse of state $|i_n\rangle$ is $|\bar\imath_n\rangle = \hat\Theta^\dagger|i_n\rangle$.

One then observes that if γ contains a D factor on the right-hand side of one of the equalities in Eq. (47), then $\overline{\gamma}$ contains the D factor on the left-hand side of the same equality, and vice versa. The weights of trajectory γ in setup A and $\overline{\gamma}$ in setup B are given by products of the factors of $D_{a\pm}^{\alpha k}$ that form each of them; this results in the *central observation* of this work [shown graphically in Fig. 5(b)],

$$\overline{P}(\overline{\gamma}) = P(\gamma)\exp[-\Delta S_{\text{res}}(\gamma)], \tag{48}$$

where $P(\gamma)$ is the weight of double trajectory γ in setup A, and $\overline{P}(\overline{\gamma})$ is that of $\overline{\gamma}$ in setup B. The reservoir entropy change, $\Delta S_{\text{res}}(\gamma)$, is the sum of the $\delta S_{\alpha k}$ for all transitions in γ.

Now consider a double trajectory γ_{d}, which goes from the n_0th state in the diagonal basis of $\rho(t_0)$ to the nth state in the diagonal basis of $\rho_{ml}^{\text{sys}}(t)$ [see Fig. 1(b)]. The subscript "d" is to indicate that it goes from diagonal basis to diagonal basis. Let us define its weight as $P(\gamma_{\text{d}})$; this equals $P(\gamma)$ multiplied by a factor of $[\mathcal{W}_0]_{i_0 n_0}[\mathcal{W}_0^\dagger]_{n_0 j_0}$ to transformation out of the diagonal basis at time t_0, and a factor of $[\mathcal{W}^\dagger]_{ni}\mathcal{W}_{jn}$ to go to the diagonal basis at time t. The unitarity of the transformations $\mathcal{W}_0$ and $\mathcal{W}$ means they do not change ΔS_{res} or ΔS_{sys}, so one has

$$\overline{P}(\overline{\gamma_{\text{d}}}) = P(\gamma_{\text{d}})\exp[-\Delta S_{\text{res}}(\gamma_{\text{d}})]. \tag{49}$$

Here, one must recall that the trajectory γ_{d} is in setup A, and goes from the state n_0 in the diagonal basis of the (initial) system density matrix at time t_0 to the state n in the diagonal basis of the (final) system's reduced density matrix at time t. Its time-reverse trajectory $\overline{\gamma_{\text{d}}}$ is a trajectory in setup B which goes from state $\overline{n}$ at time t_0 to state $\overline{n_0}$ at time t.

This relation is much the same as in Markovian stochastic thermodynamics of classical systems and which was used to derive various of the best-known fluctuation theorems [8,10,53]. In the next section, we will show that a very similar procedure allows us to derive these fluctuation theorems for quantum systems with non-Markovian dynamics. In particular, we will show that the integral fluctuation theorem in Eq. (1) holds for any system, which we will show implies that any system will satisfy the second law of thermodynamics on average, $\langle\Delta S_{\text{tot}}\rangle \geqslant 0$.

However, a complication in these quantum system (absent in the classical ones) is the question of the basis in which the dynamics are diagonal. It is crucial to note that the bases in which we consider the states $\overline{n}$ and $\overline{n_0}$ in setup B are those defined as the basis in which setup A's final and initial system density matrices are diagonal. In general, these will not both coincide with the bases in which the system density matrix for setup B will be diagonal. Consider the initial state in setup B which coincides with the time reverse of the final state in setup A; it will not evolve to a state that coincides with the initial state in setup A (cf. Sec. VII B). Thus, there is no reason to expect the final state in setup B to be diagonal in the same basis as the initial state in setup A. In this case $\overline{n_0}$ corresponds to the n_0th diagonal matrix element in the reduced system density matrix in setup B, when that density matrix is *not* written in its diagonal basis, but is written in the basis in which setup A's initial density matrix was diagonal.

This is not a problem in deriving certain fluctuation theorems, such as Eq. (1). However, our derivation of the Crooks equation only applies in those special cases in which the final state in setup B is diagonal *in the same basis* as the initial state in setup A; Sec. VIII D elaborates on this point and gives examples of special cases for which it applies.

VIII. FLUCTUATION THEOREMS

Schmiedl and Seifert showed in Ref. [8] that trajectories in classical rate equations obey Eq. (48), and one can derive most of the standard fluctuation relations from suitable sums over these classical trajectories. Their proofs were for a discrete set of states with transitions governed by Markovian rate equations. Together with a more complicated continuum version [7], this became known as *stochastic thermodynamics*, and it is discussed in a number of reviews [9,10,53]. Our objective here is to show that the same logic as in Refs. [8,10,53] can be used to derive fluctuation theorems from the Keldysh-contour trajectories using Eq. (49). Before going into the detail of the derivations for non-Markovian quantum systems, which are very close to the derivations for classical rate equations in Refs. [8,10,53], we mention the points which differ between stochastic thermodynamics for classical rate equations and for non-Markovian quantum systems.

The most obvious difference is that the trajectories themselves are very different. The quantum system's trajectories come from perturbation theory on the Keldysh contour, while the trajectories in Refs. [8,10,53] come from classical rate processes. One consequence of this is a trajectory γ_{d}, on the Keldysh contour, typically has a complex weight $P(\gamma_{\text{d}})$. However, every trajectory has a partner with the same entropy change, but with the complex conjugate weight; this trajectory is found from its partner by interchanging the trajectory's upper and lower lines and taking $\oplus \leftrightarrow \ominus$. Any physical probability will involve an equal sum of the two weights, and so will be real. None the less, this sum of a trajectory and its complex conjugate partner will often be a negative real number, so it should be considered as a contribution to the probability, and not a probability itself. The contributions with negative weights reduce the probability to go to a given state, while those with positive weights increase the probability to go to another state.

These negative weights do not occur in the usual stochastic thermodynamics of classical rate equations; however it is easy to see why. In the usual stochastic thermodynamics, the probability that a trajectory in state i has no transitions in the time window τ_n to τ_{n+1} is [8,10,53] $\exp[-\int_{\tau_n}^{\tau_{n+1}} d\tau\Gamma_i(\tau)]$, where $\Gamma_i(\tau)$ is the sum of all transition rates out of state i at time τ. To compare this with our quantum theory (which is perturbative in the reservoir couplings), such exponential terms should be expanded in powers of Γ_i. This generates a version of stochastic thermodynamics in which trajectories can have positive or negative weights. Our quantum theory has trajectories with positive and negative weights for the same reason.

The weight of a trajectory γ_{d} obeys Eq. (49); combining this with Eq. (20) gives

$$\overline{P}(\overline{\gamma_{\text{d}}})P_n(t) = P(\gamma_{\text{d}})P_{n_0}(t_0)\exp[-\Delta S_{\text{tot}}(\gamma_{\text{d}})], \tag{50}$$

where $\Delta S_{\text{tot}}(\gamma_{\text{d}})$ is the sum of the entropy change in system and reservoirs [see Eq. (16)] associated with trajectory γ_{d} from state n_0 at time t_0 to state n at time t. Despite the difference in the nature of the trajectories, this relation is the same as for classical rate equations, where t was used to derive various well-known fluctuation theorems. Now, we can follow basically the same derivations to derive the same fluctuation relations for non-Markovian quantum systems. These derivations are presented in the following subsections; readers familiar with Refs. [8,10,53] will notice their similarity to those for classical rate equations.

A. Integral fluctuation theorem

Let us start by deriving Eq. (1), which is known as the *nonequilibrium partition identity* [19–21] as well as the *integral fluctuation theorem* [8,10]. In classical systems it is the most general fluctuation theorem, since one can use stochastic thermodynamics to show that it applies to any classical system with Markovian dynamics, irrespective of that system's initial or final state. This section will show that the same is true for non-Markovian quantum systems.

If one has a physical quantity (energy, particle current, entropy, or similar) that one can calculate for each trajectory of the system, then the average value of that quantity for a system is given by the following sum over all trajectories:

$$\langle \cdots \rangle = \sum_{n_0,n} \sum_{\gamma_\mathrm{d}\in\{n_0,t_0\to n,t\}} P(\gamma_\mathrm{d}) P_{n_0}(t_0)\, (\cdots)_{\gamma_\mathrm{d}}, \tag{51}$$

where $(\cdots)_{\gamma_\mathrm{d}}$ is the quantity of interest for trajectory γ_d, and the sum is over all trajectories from state n_0 (in the diagonal basis of the system's density matrix) at time t_0 to state n (in the diagonal basis of the system's reduced density matrix) at time t.

The proof of the *integral fluctuation theorem* is carried out by considering the average,

$$\langle e^{-\Delta S_\mathrm{tot}} \rangle = \sum_{n_0,n} \sum_{\gamma_\mathrm{d}\in\{n_0,t_0\to n,t\}} P(\gamma_\mathrm{d}) P_{n_0}(t_0)\, e^{-\Delta S_\mathrm{tot}(\gamma_\mathrm{d})}, \tag{52}$$

where $P(\gamma_\mathrm{d})$ is a trajectory in the setup A defined in the paragraph above Eq. (47). Substituting in Eq. (50) on the right-hand side gives a result in terms of the trajectories in the time-reverse setup (the one called setup B above),

$$\langle e^{-\Delta S_\mathrm{tot}} \rangle = \sum_{n_0,n} \sum_{\gamma_\mathrm{d}\in\{n_0,t_0\to n,t\}} \overline{P}(\overline{\gamma_\mathrm{d}}) P_n(t). \tag{53}$$

The sum over all trajectories γ_d from n_0 at time t_0 to n at time t is replaced by a sum over all trajectories $\overline{\gamma_\mathrm{d}}$ from $\overline{n}$ at time t_0 to $\overline{n_0}$ at time t in setup B, so

$$\langle e^{-\Delta S_\mathrm{tot}} \rangle = \sum_{n_0,n} \sum_{\overline{\gamma_\mathrm{d}}\in\{\overline{n},t_0\to \overline{n_0},t\}} \overline{P}(\overline{\gamma_\mathrm{d}}) P_n(t). \tag{54}$$

Nothing changes if the sum over all n_0 is replaced by one over all $\overline{n_0}$. The dynamics of the system in setup B (whatever they may be) must conserve probability, which means that the sum over all trajectories from $\overline{n}$ to $\overline{n_0}$ summed over all $\overline{n_0}$ must give unity:

$$\sum_{\overline{n_0}} \sum_{\overline{\gamma_\mathrm{d}}\in\{\overline{n},t_0\to \overline{n_0},t\}} \overline{P}(\overline{\gamma_\mathrm{d}}) = 1. \tag{55}$$

This hold irrespective of the basis in which one writes the final state of the system, since probability conservation guarantees that the diagonal elements of a reduced density matrix sum to one in any basis. This is convenient, because the final state of the evolution in setup B (the sum over trajectories γ_d) is not usually diagonal in the basis used (which is the diagonal basis of the initial state of setup A), as discussed at the end of Sec. VII E.

Substituting Eq. (55) into Eq. (54), the right-hand side reduces to $\sum_n P_n(t)$; this is a sum over the final state of the system in setup A. However, irrespective of the dynamics of setup A, conservation of probability tells us that $\sum_n P_n(t) = 1$. Thus we have proven the *integral fluctuation theorem* in Eq. (1) under completely general conditions for an arbitrary quantum setup described by any Hamiltonian of the form Eq. (2) for any initial factorized state of system and reservoirs.

The fact the proof is restricted to factorized states of system and reservoirs means it does not apply to situations in which the system is initially entangled with reservoir states. Below, in Sec. IX, we will use the above proof as the principal ingredient in a proof of Eq. (1) for arbitrary initial states including those where the system and reservoirs are initially entangled.

However, the above proof already applies to one of the most common experimental situations, that where one has measured the system state at the beginning of the evolution in an arbitrary basis. If the basis is not the system's energy eigenbasis then the system will be in a superposition of energy states, a situation which one cannot model with the classical rate equations in Refs. [8,10,11], irrespective of whether the dynamics are Markovian or not.

B. Second law of thermodynamics

Since Eq. (1) applies for any factorizable initial state, it takes only one line of algebra [8,10,53] to arrive at the second law of thermodynamics on average:

$$\langle \Delta S_\mathrm{tot} \rangle \geqslant 0. \tag{56}$$

The proof is done by noting that $x \geqslant 1 - e^{-x}$ for all x (this is easily seen graphically, but is formally an example of Jensen's inequality), and so whatever the probability distribution of ΔS_tot, one must have $\langle \Delta S_\mathrm{tot} \rangle \geqslant 1 - \langle e^{-\Delta S_\mathrm{tot}} \rangle = 0$.

However, Eq. (1) tells us more than this; it tells us that all setups *must* sometimes have fluctuations in which $\Delta S_\mathrm{tot} < 0$. Hence, the second law is *only* obeyed on average, and there will *always* be fluctuations (perhaps only very rare fluctuations) which violate it. To see this, it is enough to note that if a setup only had trajectories with $\Delta S_\mathrm{tot}(\gamma_\mathrm{d}) > 0$ (positive entropy production), then it would have $\langle e^{-\Delta S_\mathrm{tot}} \rangle < 1$. Thus, any setup must also have trajectories with $\Delta S_\mathrm{tot}(\gamma_\mathrm{d}) < 0$ to satisfy Eq. (1). The exponential factor in Eq. (1) means that the probability of trajectories with $\Delta S_\mathrm{tot}(\gamma_\mathrm{d}) < 0$ will be less than that of those with $\Delta S_\mathrm{tot}(\gamma_\mathrm{d}) > 0$, but the probability of trajectories with $\Delta S_\mathrm{tot}(\gamma_\mathrm{d}) < 0$ cannot be zero. The only exception to this statement is a system in which no trajectories generate any entropy, so $\Delta S_\mathrm{tot}(\gamma_\mathrm{d}) = 0$ for all γ_d.

C. Jarzynski equality under certain conditions

Let us consider the Jarzynski equality [13] generalized to grand-canonical potentials [8]. It applies to a classical system that starts its evolution in thermal equilibrium at temperature T, that then experiences a time-dependent drive and time-dependent coupling to multiple reservoirs at different chemical potentials, but all at temperature T. This generalized Jarzynski equality states that the work ΔW that is done on a system by the drive and the reservoirs obeys

$$\langle e^{-\Delta W/T}\rangle = e^{-\Delta F/T}, \tag{57}$$

where temperature is measured in units of energy, so $k_{\rm B}=1$. The free energy difference

$$\Delta F = T\{\ln[Z(\mu_0;t_0)] - \ln[Z(\mu_0;t)]\}, \tag{58}$$

with $Z(\mu_0;\tau) = e^{\mu_0/T}\sum_{n_0} e^{E_{\rm sys}^{(n)}(\tau)/T}$. Here $Z(\mu_0,t_0)$ coincides with the partition function of the initial equilibrium state; however the factors of $e^{\mu_0/T}$ cancel in Eq. (58), so ΔF is independent of μ_0. The original Jarzynski equality is recovered in the limit where the system exchanges energy but not particles with the reservoirs. The above generalized Jarzynski equality was proven for classical systems described by Markovian rate equations in Refs. [8,10].

The derivation for non-Markovian quantum systems presented here is restricted to systems in which the system-reservoir coupling is reduced to zero at the end of the evolution at time t. This is in addition the assumption that system and reservoirs are in a product state (each in internal equilibrium at the same temperature T). In general, the system and reservoirs will arrive at time t in a nonfactorizable state, which will have a nonzero amount of energy in the system-reservoir coupling. Turning off the system-reservoir coupling (typically by changing the voltage on the gate that separates the system from the reservoirs) will thus change the setup's energy, and thus corresponds to work being done by the drive. We include this work done to turn off the system-environment coupling in W in Eq. (57).

Let us be clear, this restriction is a way of *avoiding* the problem of the energy in the system-reservoir coupling, by having it be zero at the beginning and end of the evolution. This is a small step beyond the proof for Markovian classical systems in Refs. [8,10], because the dynamics can be non-Markovian between time t_0 and time t (as well of course allowing for quantum physics). However, it is hoped that future work might reveal a more general Jarzynski equality, that holds when the energy in the system-reservoir coupling is nonzero at the beginning or end of the evolution. One possible direction for this future work is to compare with the situation where all reservoir chemical potentials are equal (so the reservoirs do no work on the system), for which there is an elegant proof of the Jarzynski equality in Ref. [4].

The proof presented here makes use of Eq. (48), but involves different rotations at the beginning and end of the evolution from those discussed below Eq. (48). Instead of rotations to the basis where the system's density matrix is diagonal, one rotates to the basis in which the system's Hamiltonian $H_{\rm sys}$ is diagonal. Thus $\mathcal{W}_0$ is the rotation from the diagonal basis of $H_{\rm sys}(t_0)$ to the basis in which the evolution is calculated (if these bases are the same, then $\mathcal{W}_0=1$). Similarly, $\mathcal{W}$ is the rotation from the basis in which the evolution is calculated to the basis in which $H_{\rm sys}(t)$ is diagonal. While these rotations are different from those below Eq. (48), they are still unitary, which means they do not affect the trajectory's entropy; hence Eq. (49) still holds.

Consider a trajectory $\gamma_{\rm d}$ from system state n_0 at time t_0 to system state n in time t, where n_0 is the eigenstate of $H_{\rm sys}(t_0)$ with energy $E_{n_0}(t_0)$ and n is the eigenstate of $H_{\rm sys}(t)$ with energy $E_n(t)$. Then the work done on the system by the driving is

$$\Delta W_{\rm drive}(\gamma_{\rm d}) = \left[E_{\rm sys}^{(n)}(t) - E_{\rm sys}^{(n_0)}(t_0)\right] + \sum_\alpha \Delta E_\alpha(\gamma_{\rm d}). \tag{59}$$

The square brackets give the work done by the drive which stays in the system, while $\Delta E_\alpha(\gamma_{\rm d})$ is defined as the energy flow into reservoir α during the trajectory $\gamma_{\rm d}$. Note that this equality holds because of the above restriction to the system in which there is no energy in the system-reservoir coupling at the beginning or end of the evolution. The work done on the system by the reservoirs during trajectory $\gamma_{\rm d}$ is given by

$$\Delta W_{\rm res}(\gamma_{\rm d}) = -\sum_\alpha \mu_\alpha \Delta N_\alpha(\gamma_{\rm d}), \tag{60}$$

where $\Delta N_\alpha(\gamma_{\rm d})$ is the number of particles flowing into reservoir α during the trajectory $\gamma_{\rm d}$. Given Eq. (11), one sees that

$$\Delta S_{\rm res}(\gamma_{\rm d}) = \frac{1}{T}\sum_\alpha [\Delta E_\alpha(\gamma_{\rm d}) - \mu_\alpha \Delta N_\alpha(\gamma_{\rm d})], \tag{61}$$

since all reservoirs have the same temperature. Thus,

$$\Delta W(\gamma_{\rm d}) = E_{\rm sys}^{(n)}(t) - E_{\rm sys}^{(n_0)}(t_0) + T\Delta S_{\rm res}(\gamma_{\rm d}), \tag{62}$$

where $\Delta W(\gamma_{\rm d}) = \Delta W_{\rm drive}(\gamma_{\rm d}) + \Delta W_{\rm res}(\gamma_{\rm d})$ is the total work done on the system. Using this equality in the average of $\exp[-\Delta W(\gamma_{\rm d})/T]$ over all $\gamma_{\rm d}$, defined in Eq. (51),

$$\begin{aligned}\langle e^{-\Delta W/T}\rangle = \sum_{n_0,n}\ \sum_{\gamma_{\rm d}\in\{n_0,t_0\to n,t\}} & P(\gamma_{\rm d}) e^{-\Delta S_{\rm res}(\gamma_{\rm d})} \\ &\times e^{-[E_{\rm sys}^{(n)}(t) - E_{\rm sys}^{(n_0)}(t_0)]/T} P_{n_0}(t_0),\end{aligned} \tag{63}$$

Eq. (48) is now used to write this in terms of $\overline{P}(\overline{\gamma_{\rm d}})$. In other words, the average over trajectories in a setup A is written in terms of the trajectories in setup B (defined earlier as the time reverse of setup A). The initial system density matrix (at time t_0) is diagonal in the eigenbasis of $H_{\rm sys}(t_0)$, and the probability of being in state n_0 is

$$P_{n_0}(t_0) = \frac{1}{Z_0(\mu_0)}\, e^{-[E_{\rm sys}^{(n_0)}(t_0) - \mu_0]/T}, \tag{64}$$

with $Z_0(\mu_0)$ given below Eq. (57). Then Eq. (63) becomes

$$\langle e^{-\Delta W/T}\rangle = \frac{1}{Z_0(\mu_0)}\sum_{n_0,n} e^{-[E_{\rm sys}^{(n)}(t) - \mu_0]/T} \sum_{\overline{\gamma_{\rm d}}\in\{\overline{n},t_0\to\overline{n_0},t\}} \overline{P}(\overline{\gamma_{\rm d}}), \tag{65}$$

where the sum over all $\gamma_{\rm d}$ from n_0 to n in setup A has become a sum over all $\overline{\gamma_{\rm d}}$ from $\overline{n}$ to $\overline{n_0}$ in setup B. Nothing changes if the sum over all n_0 is replaced by one over all $\overline{n_0}$. Irrespective of the dynamics in setup B, the sum over all trajectories with final state $\overline{n_0}$, summed over all $\overline{n_0}$, must give one. Therefore Eq. (65) reduces to $\langle e^{-\Delta W/T}\rangle = Z(\mu_0)/Z_0(\mu_0)$. Now using Eq. (58), one immediately gets the generalized Jarzynski equality in Eq. (57).

One can apply Jensen's inequality to Eq. (57) to find a well-known formulation of the second law,

$$\langle \Delta W\rangle \leqslant \Delta F, \tag{66}$$

but the assumptions and restrictions in this derivation to Eq. (57) make this a less general version of the second law than that in Eq. (56).

D. Crooks equation

The Crooks equation[15] is a relation between the dynamics of a setup A and a setup B (which has the time reverse of setup A) in situations where the system undergoes time-dependent driving while coupled to reservoirs. Consider setup A described by the Hamiltonian in Eq. (2) starting at time t_0 with the system's density matrix $\rho_{\rm sys}^{\rm (i)}$ ("i" for initial), and ending the evolution at time t with the systems reduced density matrix being $\rho_{\rm sys}^{\rm (f)}$ ("f" for final). Let us define $P(\Delta S_{\rm tot};\rho_{\rm sys}^{\rm (i)}\to\rho_{\rm sys}^{\rm (f)})$ as the probability that setup A would have a total entropy changes of $\Delta S_{\rm tot}$ between t_0 and t. Now let us consider setup B described by the time reverse of Eq. (2), and take its initial system density matrix to be $\overline{\rho}_{\rm sys}^{\rm (f)}\equiv\Theta_{\rm sys}^{\dagger}\rho_{\rm sys}^{\rm (f)}\Theta_{\rm sys}$ where $\Theta_{\rm sys}$ is the time-reversal operator on the system alone; so setup B's initial system state is the time reverse of setup A's final system state. Setup B's evolution will not be the time reverse of setup A's, because we do not time-reverse individual reservoir states; cf. Sec. VII B. Let its evolution be under the time reverse of Eq. (2), so that its reduced system density matrix at time t is $\overline{\rho}_{\rm sys}^{\rm (f2)}$. Let us then define $\overline{P}(\Delta S_{\rm tot};\overline{\rho}_{\rm sys}^{\rm (f)}\to\overline{\rho}_{\rm sys}^{\rm (f2)})$ as the probability that setup B would have a total entropy changes of $\Delta S_{\rm tot}$ between t_0 and t. Below we will prove the following slight generalization of the Crooks equation for non-Markovian quantum systems; it reads

$$\overline{P}\big(-\Delta S_{\rm tot};\overline{\rho}_{\rm sys}^{\rm (f)}\to\overline{\rho}_{\rm sys}^{\rm (f2)}\big)=P\big(\Delta S_{\rm tot};\rho_{\rm sys}^{\rm (i)}\to,\rho_{\rm sys}^{\rm (f)}\big)e^{-\Delta S_{\rm tot}}, \tag{67}$$

under the condition that the time reverse of the final state of the evolution in setup B, $\rho_{\rm sys}^{\rm (f2)}$, is diagonal in the same basis as the initial state in setup A, $\rho_{\rm sys}^{\rm (i)}$. This is slightly more general than the condition used by Crooks for classical systems [15], since his condition was that the final state in setup B was the same as (the time reverse [55] of) the initial state in setup A.

In general, there is no reason to expect the condition below Eq. (67) to hold; if setup B's initial state is the time reverse of the final state of setup A, it is likely to end up in some state $\overline{\rho}_{\rm sys}^{\rm (f2)}$, whose time reverse $\rho_{\rm sys}^{\rm (f2)}$ has nothing to do with $\rho_{\rm sys}^{\rm (i)}$. Thus, in general Eq. (67) will not be satisfied, but there are scenarios of interest in which the condition is satisfied. Figure 6 shows a situation (a quantum version of a scenario proposed by Crooks [15]) in which one naturally has $\rho_{\rm sys}^{\rm (f2)}=\rho_{\rm sys}^{\rm (i)}$.

One can easily generalize the scenario in Fig. 6 to one where $\rho_{\rm sys}^{\rm (f2)}$ is diagonal in the same basis as $\rho_{\rm sys}^{\rm (i)}$ without equaling $\rho_{\rm sys}^{\rm (i)}$. This generalization is one where setup A's dynamics between time t_0 and t_1 (and hence setup B's dynamics between time $\overline{t_1}$ and t) involve strong Markovian decoherence, but need not have relaxation. Let us assume that setup A starts with a nonequilibrium $\rho_{\rm sys}^{\rm (i)}$ which is diagonal in the diagonal basis of $H_{\rm sys}(t_0)$. Then at time t_1 it will still be diagonal in that basis, but after that its evolution will generate an arbitrary state, $\rho_{\rm sys}^{\rm (f)}$, at time t. Taking the initial state in setup B as $\overline{\rho}_{\rm sys}^{\rm (f)}$, the dynamics of setup B will give some other state at time $\overline{t_1}$ (the red square in Fig. 6). However, this state will be completely decohered between time $\overline{t_1}$ and time t, irrespective of whether it relaxes or not. This means that setup B's reduced system density matrix at time t will be diagonal in the diagonal basis of setup B's Hamiltonian at time t. Thus the time reverse of this state will be diagonal in the same basis as $\rho_{\rm sys}^{\rm (i)}$. Hence, it satisfies the condition below Eq. (67), even though the state $\rho_{\rm sys}^{\rm (f2)}\neq\rho_{\rm sys}^{\rm (i)}$ and $\rho_{\rm sys}^{\rm (f2)}$ may be very far from equilibrium (if little or no relaxation has occurred).

Proof of the Crooks equation

To derive Eq. (67) from Eq. (48) we can follow the proof in Ref. [10]. The probability that the entropy change is $\Delta S_{\rm tot}$ in the time from t_0 to t is

$$P\big(\Delta S_{\rm tot};\rho_{\rm sys}^{\rm (i)}\to\rho_{\rm sys}^{\rm (f)}\big)=\sum_{n_0,n}\ \sum_{\gamma_{\rm d}\in\{n_0,t_0\to n,t\}}P(\gamma_{\rm d})P_{n_0}^{\rm (i)}\times\delta[\Delta S_{\rm tot}(\gamma_{\rm d})-\Delta S_{\rm tot}], \tag{68}$$

where the δ function picks out only those trajectories with entropy change $\Delta S_{\rm tot}$, and $P_{n_0}^{\rm (i)}$ is the n_0th element of $\rho_{\rm sys}^{\rm (i)}$ in its diagonal basis. The δ function means that the equality holds if one multiples the left-hand side by $e^{-\Delta S_{\rm tot}}$ and the right-hand side by $e^{-\Delta S_{\rm tot}(\gamma_{\rm d})}$. Equation (50)—derived above from Eq. (48)—can be used to write

$$P\big(\Delta S_{\rm tot};\rho_{\rm sys}^{\rm (i)}\to\rho_{\rm sys}^{\rm (f)}\big)e^{-\Delta S_{\rm tot}}=\sum_{n_0,n}\ \sum_{\gamma_{\rm d}\in\{n_0,t_0\to n,t\}}\overline{P}(\overline{\gamma_{\rm d}})P_n^{\rm (f)}\delta[\Delta S_{\rm tot}(\gamma)-\Delta S_{\rm tot}].$$

This means the dynamics are now written in terms of trajectories in the time-reversed setup (setup B). Rewriting the sum over $\gamma_{\rm d}$ from n_0 to n as a sum over $\overline{\gamma_{\rm d}}$ from $\overline{n}$ to $\overline{n_0}$, and using the fact that $\overline{\Delta S_{\rm tot}(\gamma)}=-\Delta S_{\rm tot}(\gamma)$, leads to

$$P\big(\Delta S_{\rm tot};\rho_{\rm sys}^{\rm (i)}\to\rho_{\rm sys}^{\rm (f)}\big)e^{-\Delta S_{\rm tot}}=\sum_{\overline{n_0},\overline{n}}\ \sum_{\overline{\gamma_{\rm d}}\in\{\overline{n},t_0\to\overline{n_0},t\}}\overline{P}(\overline{\gamma_{\rm d}})P_n^{\rm (f)}\delta[\Delta S_{\rm tot}(\gamma)+\Delta S_{\rm tot}], \tag{69}$$

where we have used the fact that nothing changes when the sum over n_0 and n is replaced by a sum over $\overline{n_0}$ and $\overline{n}$. Recalling that $P_n^{\rm (f)}$ is the probability that setup A finishes in state n in the diagonal basis of $\rho_{\rm sys}^{\rm (f)}$, one can always choose the initial density matrix in setup B to be $\overline{\rho}_{\rm sys}^{\rm (f)}$. Then, the probability that the system starts in state $\overline{n}$ in the diagonal basis of $\overline{\rho}_{\rm sys}^{\rm (f)}$ equals $P_n^{\rm (f)}$. Hence, it looks like the right-hand side of Eq. (69) equals $\overline{P}(-\Delta S_{\rm tot};\overline{\rho}_{\rm sys}^{\rm (f)}\to\overline{\rho}_{\rm sys}^{\rm (f2)})$, whatever final system density matrix, $\overline{\rho}_{\rm sys}^{\rm (f2)}$, this evolution may give. However, this is overlooking the fact that the trajectories $\overline{\gamma_{\rm d}}$ end in the diagonal basis of $\overline{\rho}_{\rm sys}^{\rm (i)}$; thus the right-hand side of Eq. (69) only equals $\overline{P}(-\Delta S_{\rm tot};\overline{\rho}_{\rm sys}^{\rm (f)}\to\overline{\rho}_{\rm sys}^{\rm (f2)})$, if $\rho_{\rm sys}^{\rm (f2)}$ is diagonal in the same basis as $\rho_{\rm sys}^{\rm (i)}$. So one only recovers Eq. (67) if the dynamics satisfy the condition below Eq. (67).

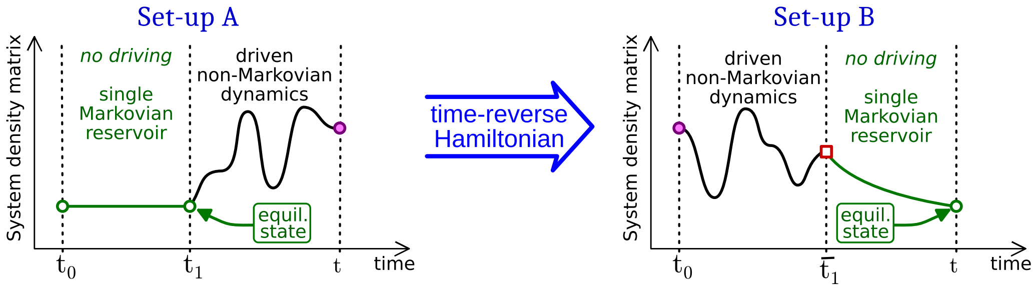

FIG. 6. A sketch of a situation for which the Crooks equation in Eq. (67) is applicable. The plots show a cartoon of how the system density matrix (vertical axis) varies with time from t_0 to t. In setup A, the system starts in a thermal state (open green circle). For time t_0 to time t_1, the system has a coupling to a single Markovian reservoir with which it is in equilibrium. Thus at time t_1 the system is still in the thermal state. For time t_1 to time t, the system is driven while interacting with many non-Markovian reservoirs, so that at time t it is in a highly nonequilibrium state (filled purple circle). In setup B the system starts in the time reverse of setup A's final state, and evolves under the time-reversed Hamiltonian. Between t_0 and $\overline{t_1} = t + t_0 - t_1$ it undergoes driven non-Markovian dynamics, so it is in some nonequilibrium state (red square) at time $\overline{t_1}$. However, after that it is only coupled to a single Markovian reservoir, so it decays towards the thermal state in equilibrium with that reservoir. We assume it reaches that state at time t. Thus the initial state in both setups is the final state in the other, which is sufficient that Eq. (67) is applicable, even though the evolution in setup A is completely arbitrary and non-Markovian between t_1 and t. Finally, as nothing happens to the state in setup A between t_0 and t_1, we can equally find a Crooks equation of the type in Eq. (67) between the dynamics from time t_1 to time t in setup A and the dynamics from time t_0 to time t in setup B.

IX. FLUCTUATION THEOREMS FOR NONFACTORIZABLE INITIAL CONDITIONS

The system and reservoirs can be in either a factorizable or nonfactorizable state, a factorizable state being one where the total density matrix can be written as a product of the system density matrix and the reservoir density matrices, e.g., $\hat{\rho}_{\rm sys} \otimes \hat{\rho}_{\rm res1} \otimes \hat{\rho}_{\rm res2} \otimes \cdots$. This state is often also called a product state. A nonfactorizable state is any other density matrix for the system plus the reservoirs. Up to this point, this work has discussed a setup which started its evolution at time t_0 in a factorized state. In this section, we consider protocols in which the initial density matrix is in a nonfactorizable state.

In quantum mechanical systems, a system's state is changed by the mere fact of observing it. In particular, the act of measuring the system state projects it into a definite system state, which means the entanglement with the reservoirs is destroyed, leaving the system and reservoirs in a factorized density matrix. Hence, the only way to measure the changes between time t_1 and time t_2 without this projection onto a factorized state at time t_1 is to consider the following protocol.

Nonfactorizing protocol. We prepare many setups in the same manner (starting each with the same factorized state at a time t_0 and letting them evolve in the same manner), so they are all in the same nonfactorized state at a time t_1. We then split them into two groups (i and ii). We measure group i at time t_1 and we measure group ii at a later time t_2. As we do not measure the setups in group ii at time t_1, they are not projected onto a factorized state at time t_1. Despite this, we know about the state of the system at time t_1 from the measurements on setups in group i. This enables us to see the difference between the setup's properties at time t_2, and its properties at the earlier time t_1, when it was in a nonfactorized state at time t_1.

It is important to note that this protocol cannot be used to study correlations between the state at time t_1 and time t_2, because one measurement is on group i and the other is on group ii. For example, we can see how the distribution of entropy changes between time t_1 and time t_2, but we cannot see how a fluctuation of entropy (say the system having much less entropy than average) at time t_1 correlates with a fluctuation of entropy at the later time t_2.

A. Conditional probability in this protocol

Let us consider the above nonfactorizing protocol being used to study the changes in a setup between time t_1 and time t_2, when the system is in a nonfactorized state at time t_1. For this one can assume the setup was prepared in the distant past at time t_0 in a factorized state, but that the system has interacted with the reservoirs for so long by the time t_1 that it is in a highly complicated entangled state with the reservoirs. Our main interest is in situations where the time t_0 was so far in the past that the dynamics at the times of interest (t_1 and t_2) *do not depend* on the choice of the system state at time t_0.

Consider $P(\Delta S_1^{\rm tot}; t_1, t_0)$ to be the probability distribution of the entropy change $\Delta S_1^{\rm tot}$ between the time in the distant past t_0 and time t_1, as measured on setups in group i. Then consider $P(\Delta S_2^{\rm tot}; t_2, t_0)$ to be the probability distribution of the entropy change $\Delta S_2^{\rm tot}$ between the time in the distant past t_0 and time t_2, as measured on setups in group ii. Then one can define $Q(\Delta S_{2\leftarrow 1}^{\rm tot}; t_2, t_1)$ as a conditional probability distribution, for the entropy change of

$$\Delta S_{2\leftarrow 1}^{\rm tot} = \Delta S_2^{\rm tot} - \Delta S_1^{\rm tot}, \tag{70}$$

between time t_1 and time t_2. This means that $Q(\Delta S_{2\leftarrow 1}^{\rm tot}; t_2, t_1)$ measures how the probability distribution changes between t_1 and t_2. It obeys

$$P\big(\Delta S_2^{\rm tot}; t_2, t_0\big) = \int d\big(\Delta S_1^{\rm tot}\big)\, Q\big(\Delta S_2^{\rm tot} - \Delta S_1^{\rm tot}; t_2, t_1\big) \times P\big(\Delta S_1^{\rm tot}; t_1, t_0\big). \tag{71}$$

One can always define the function $Q(\Delta S_{2\leftarrow 1}^{\rm tot}; t_2, t_1)$ in this manner. However, the price to pay for highly non-Markovian dynamics (strong memory effects) is that it may depend on both the initial state of the setup at time t_0 and on the dynamics of the setup from time t_0 to time t_1 (as well as the dynamics from t_1 to t_2). Thus this is not a pleasant quantity to consider in general. However, it becomes much more natural in situations where t_0 is far enough in the past that $Q(\Delta S^{\rm tot}_{2\leftarrow 1};t_2,t_1)$ depends weakly on it, and where the dynamics for a long time before time t_1 are simple enough to treat in some manner. An ideal example, which we will consider in more detail below, is when the system Hamiltonian is time-independent for a long enough time before t_1 that the setup has achieved a *steady state* at time t_1.

B. Integral fluctuation theorem

Now we use the proof of the integral fluctuation theorem, Eq. (1), for factorized initial conditions, to prove that it also holds for the entropy change between time t_1 and t_2, when the setup is in an arbitrary nonfactorized state at both t_1 and t_2. In this context, we assume the entropy change is measured via the nonfactorizing protocol above, in which the setup was in a factorized state at a time t_0 in the distant past (long before the times of interest, t_1 and t_2).

As above we assume $\Delta S_2^{\rm tot}$ is the entropy change from time t_0 to time t_2 (as measured on setups in group ii of the nonfactorizing protocol). Then Eq. (1), proven for a factorized state at time t_0 in Sec. VIII A above, becomes

$$\langle e^{-\Delta S_2^{\rm tot}}\rangle = \int d\big(\Delta S_2^{\rm tot}\big)P\big(\Delta S_2^{\rm tot};t_2,t_0\big)e^{-\Delta S_2^{\rm tot}}. \tag{72}$$

Substituting in Eq. (71),

$$\begin{aligned}\langle e^{-\Delta S_2^{\rm tot}}\rangle &= \int d\big(\Delta S^{\rm tot}_{2\leftarrow 1}\big)d\big(\Delta S_1^{\rm tot}\big)\ Q\big(\Delta S^{\rm tot}_{2\leftarrow 1};t_2,t_1\big)\\ &\quad\times P\big(\Delta S_1^{\rm tot};t_1,t_0\big)e^{-\Delta S^{\rm tot}_{2\leftarrow 1}-\Delta S_1^{\rm tot}}\\ &= \int d\big(\Delta S^{\rm tot}_{2\leftarrow 1}\big)Q\big(\Delta S^{\rm tot}_{2\leftarrow 1};t_2,t_1\big)e^{-\Delta S^{\rm tot}_{2\leftarrow 1}}\langle e^{-\Delta S_1^{\rm tot}}\rangle,\end{aligned} \tag{73}$$

where $\langle\exp[-\Delta S_1^{\rm tot}]\rangle$ is the average over dynamics from time t_0 to t_1. Now substituting in Eq. (1) for the two averages, we have

$$1 = \int d\big(\Delta S^{\rm tot}_{2\leftarrow 1}\big)\ Q\big(\Delta S^{\rm tot}_{2\leftarrow 1};t_2,t_1\big)e^{-\Delta S^{\rm tot}_{2\leftarrow 1}} \equiv \langle e^{-\Delta S^{\rm tot}_{2\leftarrow 1}}\rangle, \tag{74}$$

where $\Delta S^{\rm tot}_{2\leftarrow 1}$ is the entropy change in the setup between time t_1 and t_2. Hence, we have the integral fluctuation theorem in Eq. (1) for any nonfactorized initial state, the initial state now being the time at which one starts to study the setup (time t_1).

The average, $\langle\cdots\rangle$, in Eq. (74) is defined via the nonfactorizing protocol above, which relates changes to the difference between the setup's nonfactorized state at time t_1 (as measured on setups in group i of the nonfactorizing protocol) and the setup's nonfactorized state at a later time t_2 (as measured on setups in group ii of the nonfactorizing protocol).

It immediately follows from this proof that all statements about the second law of thermodynamics in Sec. VIII B above also hold for nonfactorized states. The second law is always true on average,

$$\big\langle\Delta S^{\rm tot}_{2\leftarrow 1}\big\rangle \geqslant 0, \tag{75}$$

irrespective of whether the setup is in a factorized state at time t_1 or not. Hence the average entropy will never be smaller at time t_2 than at time t_1 (for any $t_2 > t_1$), However, there *must* also be fluctuations for which $\Delta S^{\rm tot}_{2\leftarrow 1} < 0$, if Eq. (74) is to be satisfied.

C. Steady-state fluctuation relation

We can expect that a large class of non-Markovian systems will decay to a situation of steady state flow, if the system is coupled to two or more reservoirs at different temperatures and electrochemical potentials, while the Hamiltonian is kept time-independent. Here we consider the case where there is a single steady state (for a given Hamiltonian and given reservoir parameters), which *all initial states decay to*. This steady state will generally not be a factorizable state of the system and the reservoirs, since the system will be entangled with at least some reservoir modes at all times.

The objective here is to derive the Evans-Searles fluctuation relation [14] for such a non-Markovian system for which the steady state is nonfactorizable. For this, consider the nonfactorizing protocol above, in which time t_0 is so far in the past that the choice of initial state at t_0 is irrelevant for the steady-state dynamics at the times of interest (t_1 and t_2). Since one is completely free to choose the system state at time t_0, take it to coincide with that given by the steady state when one traces out the reservoirs. Then, by construction, the initial-system density matrix and the reduced final-system density matrix are the same. This means the setup obeys the Crooks equality derived above in Sec. VIII D. We also assume the Hamiltonian is invariant under time reversal, such as is the case if it is time-independent, and has no external magnetic field. This means that the dynamics in the time-reversed setup are the same as in the original setup. Then the Crooks equality for the entropy change between time t_0 and time t_2 reads

$$P\big(-\Delta S_2^{\rm tot};t_2,t_0\big) = P\big(\Delta S_2^{\rm tot};t_2,t_0\big)e^{-\Delta S_2^{\rm tot}}, \tag{76}$$

where we have dropped the overline on the left, because time reversal changes nothing. Now Eq. (71) is used to write the right-hand side as evolution from time t_0 to time t_1 followed by evolution from t_1 to t_2, as follows:

$$\begin{aligned}&P\big(\Delta S_2^{\rm tot};t_2,t_0\big)e^{-\Delta S_2^{\rm tot}}\\ &\quad= e^{-\Delta S_2^{\rm tot}}\int d\big(\Delta S_1^{\rm tot}\big)Q\big(\Delta S_2^{\rm tot}\\ &\qquad -\Delta S_1^{\rm tot};t_2,t_1\big)P\big(\Delta S_1^{\rm tot};t_1,t_0\big).\end{aligned} \tag{77}$$

By the same logic the left-hand side of Eq. (76) is

$$\begin{aligned}&P\big(-\Delta S_2^{\rm tot};t_2,t_0\big)\\ &\quad= \int d\big(\Delta S_1^{\rm tot}\big)Q\big(\Delta S_1^{\rm tot}\\ &\qquad -\Delta S_2^{\rm tot};t_2,t_1\big)P\big(-\Delta S_1^{\rm tot};t_1,t_0\big)\\ &\quad= e^{-\Delta S_2^{\rm tot}}\int d(\Delta S_1^{\rm tot})Q\big(\Delta S_1^{\rm tot}-\Delta S_2^{\rm tot};t_2,t_1\big)\\ &\quad\times e^{\Delta S_2^{\rm tot}-\Delta S_1^{\rm tot}}P\big(\Delta S_1^{\rm tot};t_1,t_0\big),\end{aligned} \tag{78}$$

where the last line comes from substituting in the Crooks equality, as applied to the evolution from time t_0 to t_1, for which it takes the form $P(-\Delta S_1^{\rm tot};t_1,t_0)=P(\Delta S_1^{\rm tot};t_1,t_0)e^{-\Delta S_1^{\rm tot}}$.

Note that the integrals in Eq. (77) and Eq. (78) are both convolutions of $P(\Delta S_1^{\rm tot};t_1,t_0)$ with another function; in the former case that function is $Q(\Delta S_2^{\rm tot}-\Delta S_1^{\rm tot};t_2,t_1)$ and in the latter case that function is $Q(-\Delta S_2^{\rm tot}+\Delta S_1^{\rm tot};t_2,t_1)e^{\Delta S_2^{\rm tot}-\Delta S_1^{\rm tot}}$. Substituting Eq. (77) and Eq. (78) into the right- and left-hand sides of Eq. (76) gives us an equality between the two convolutions,

$$\begin{aligned}&\int d\big(\Delta S_1^{\rm tot}\big)Q\big(\Delta S_2^{\rm tot}-\Delta S_1^{\rm tot};t_2,t_1\big)P\big(\Delta S_1^{\rm tot};t_1,t_0\big)\\ &\quad=\int d\big(\Delta S_1^{\rm tot}\big)Q\big(\Delta S_1^{\rm tot}-\Delta S_2^{\rm tot};t_2,t_1\big)\\ &\qquad\times e^{\Delta S_2^{\rm tot}-\Delta S_1^{\rm tot}}P\big(\Delta S_1^{\rm tot};t_1,t_0\big).\end{aligned}\tag{79}$$

This equality has the mathematical structure

$$\int dx A_1(y-x)B(x)=\int dx A_2(y-x)B(x)\quad\text{for all } y.$$

We wish to show that this implies that the functions $A_1(x)$ and $A_2(x)$ are identical, irrespective of the form of $A_1(x)$ and $B(x)$. To do this we consider the Fourier transforms of the functions defined as $A_i(x)=\int dk a_i(k)e^{ikx}$ for $i=1,2$, and $B(x)=\int dk b(k)e^{ikx}$. We assume that the functions $a_1(k)$, $a_2(k)$, and $b(k)$ are well behaved, and also assume that $b(k)$ is not zero over any finite range of k. Given the Fourier transforms we have

$$\int dy\, e^{-iky}\int dx\, A_1(y-x)B(x)=(2\pi)^2a_1(k)b(k),\tag{80}$$

with a similar equation for A_2 in place of A_1. This immediately gives $a_2(k)=a_1(k)$ for all k where $b(k)\neq 0$ [however it tells us nothing about the relationship between $a_2(k)$ and $a_1(k)$ when $b(k)=0$]. So long as $b(k)$ is not zero over a finite range of k, then it is sufficient to perform the inverse Fourier transform on $a_1(k)$ and $a_2(k)$, to find $A_2(x)=A_1(x)$ for all x.

This means that so long as the Fourier transform of all the probability distributions in Eq. (79) are well behaved and the Fourier transform of $P(\Delta S_1^{\rm tot};t_1,t_0)$ only vanishes at discrete points, then

$$Q\big(-\Delta S_{2\leftarrow 1}^{\rm tot};t_2,t_1\big)=Q\big(\Delta S_{2\leftarrow 1}^{\rm tot};t_2,t_1\big)e^{-\Delta S_{2\leftarrow 1}^{\rm tot}},\tag{81}$$

where $\Delta S_{2\leftarrow 1}^{\rm tot}$ is the entropy change between time t_1 and time t_2, given by Eq. (70). This is the Evans-Searles steady-state fluctuation relation derived for a nonfactorizable steady state in a non-Markovian system. The derivation holds for any situation where all initial system states decay to the same steady state.

A careful reader will note that the proof is a little more general; it can also hold for a system with multiple steady states (A, B, etc.), so long as the initial factorized state with the system density matrix which corresponds to the reduced density matrix of steady state A does indeed decay to steady state A (and not to steady state B). This is plausible, since one would imagine that this initial state is the closest product state to steady state A, but there may be systems that violate it. Of course the proof does not apply to systems which do not decay to steady states, such as those that decay to limit cycles.

X. APPROXIMATE THEORIES

This work connects fluctuation theorems to a microscopic symmetry of the system-reservoirs interactions, going beyond Ref. [4]. This can be used to identify a family of approximations which are guaranteed to satisfy fluctuation theorems. These approximations must contain a trajectory $\overline{\gamma}$ for every trajectory γ, and individual transitions must satisfy local-detailed balance, thereby satisfying Eqs. (47). Then the above arguments apply, so Eq. (48) is recovered, which leads to all the usual fluctuation theorems, which means they will always obey the second law on average.

The first approximation is the Born approximation for weak system-reservoir coupling, also called the Bloch-Redfield [56,57] or sequential tunneling approximation [30]; see also Refs. [44,45,58,59] or various textbooks [60–62]. This neglects trajectories where the system interacts with multiple reservoir modes at the same time, which is reasonable when the coupling is weak on the scale of the reservoir's memory time. The approximation has a trajectory $\overline{\gamma}$ for every γ, and individual transitions satisfy local-detailed balance, which is enough to prove that it obeys all the usual fluctuation theorems. For strictly vanishing memory time (Markovian dynamics), this reduces to a Lindblad equation [44,63,64], for which a different proof of fluctuation theorems exists [46]. However, our proof applies equally to systems with short (but nonzero) memory times.

Next is the cotunneling approximation [30], in which the system can interact with two reservoir modes at the same time. This is a used in Coulomb-blockaded quantum dots, where it can dominate the transport in certain regimes [30]. Since this approximation obeys the conditions discussed above, this constitutes a proof that the cotunneling approximation obeys all the usual fluctuation theorems. Similarly, by allowing up to n simultaneous interactions with reservoir modes (for different n), one gets a family of approximations which all obey the fluctuation theorems.

XI. CONCLUSIONS

This work uses a real-time diagrammatic theory on the Keldysh contour to develop the *quantum stochastic thermodynamics* of arbitrary systems coupled to ideal reservoirs. It shows that energy conservation ensures that the system obeys the first law of thermodynamics on average. Then, by finding the symmetry between trajectories on the Keldysh contour in Eq. (48), it shows that the integral fluctuation theorem, Eq. (1), holds for all non-Markovian system dynamics, including nonfactorized initial conditions, so these dynamics obey the second law on average. It gives other fluctuation theorems, such as Jarzynski or Crooks, in the right conditions. Similarly, a nonfactorized steady state obeys the Evans-Searles fluctuation relation [14], if the Hamiltonian in Eq. (2) is invariant under time reversal.

The most obvious practical consequence of these results is that they prove that no quantum machine (Markovian or non-Markovian) will ever exceed Carnot efficiency on average.

A family of approximations is identified which satisfies Eq. (48) and so fulfills the fluctuation theorems. This provides a powerful tool to analyze nanoscale energy harvesting and refrigeration beyond weak coupling.

ACKNOWLEDGMENTS

I thank A. Auffeves, D. Basko, M. Campisi, A. Crepieux, C. Elouard, M. Esposito, E. Jussiau, and F. Michelini for useful comments or discussions. I acknowledge the financial support of the COST Action MP1209 "Thermodynamics in the quantum regime", the CNRS PEPS grant "ICARE", and the French National Research Agency's "Investissement d'avenir" program (ANR-15-IDEX-02) via the Université Grenoble Alpes QuEnG project.

APPENDIX: REMINDER ON TIME REVERSAL IN QUANTUM MECHANICS

Here we recall the results that we will need related to time reversal in quantum mechanics, which can be found in Messiah's famous textbook [54]. First, the time inversion of a quantum state $|i\rangle$ is defined as

$$|\bar{i}\rangle = \hat{\Theta}^\dagger |i\rangle, \tag{A1}$$

where $\hat{\Theta}^\dagger$ is the time-inversion operator. In the absence of spins, time inversion of a wave function is just taking its complex conjugate; thus $\hat{\Theta}^\dagger = \hat{\Theta}_0^\dagger$, where $\hat{\Theta}_0^\dagger$ is the complex-conjugation operator. To understand the role of $\hat{\Theta}_0^\dagger$ for a single-particle problem, one notes that position states are invariant under time inversion, and so if one writes the system wave function $|i\rangle$ as a vector of position states, then $\hat{\Theta}_0^\dagger$ is the operator which takes the complex conjugate of all elements of the vector. For a many-body problem, the same is also true, if one writes the system state as a vector of many-body position states (with a position for each particle). Defining $\hat{\Theta}_0$ such that $\hat{\Theta}_0\hat{\Theta}_0^\dagger = 1$, one has $\hat{\Theta}_0^\dagger \mathcal{X} \hat{\Theta}_0 = \mathcal{X}^*$ for any matrix $\mathcal{X}$ written in a basis of many-body position states.

In the presence of spin-halves, the time-inversion operator also flips the spins about the y axis, so

$$\hat{\Theta}^\dagger = -i\sigma_y \hat{\Theta}_0^\dagger. \tag{A2}$$

The time inversions of a position operator, $\hat{x}$, a momentum operator, $\hat{p} = -i\hbar\, d/dx$, and a Pauli spin operator $\hat{\sigma}_\alpha$ are

$$\overline{\hat{x}} = \hat{\Theta}^\dagger \hat{x} \hat{\Theta} = \hat{x}, \tag{A3a}$$

$$\overline{\hat{p}} = \hat{\Theta}^\dagger \hat{p} \hat{\Theta} = -\hat{p}, \tag{A3b}$$

$$\overline{\hat{\sigma}}_\alpha = \hat{\Theta}^\dagger \hat{\sigma}_\alpha \hat{\Theta} = -\hat{\sigma}_\alpha. \tag{A3c}$$

The time reverse of a Hamiltonian in the time window t_0 to t as sketched in Fig. 4 is

$$\begin{aligned}\overline{\hat{H}(B,\sigma_\alpha,\tau)} &= \hat{\Theta}^\dagger \hat{H}(B,\hat{\sigma}_\alpha, t_0+t-\tau)\hat{\Theta} \\ &= \hat{H}(-B,-\hat{\sigma}_\alpha, t_0+t-\tau),\end{aligned} \tag{A4}$$

where the dependence of $\mathcal{H}$ on external fields, B, and Pauli spin matrices, σ_α, is explicitly shown to recall how they transform under time reversal. The evolution operator from time t_0 to time τ under such a time-dependent Hamiltonian [the solid part of the curve in Fig. 4(a)] is given by the usual time-ordered integral

$$\hat{U}(\tau;t_0) = \mathcal{T}\exp\left[-i\int_{t_0}^{\tau} d\tau' \hat{H}(\tau')\right], \tag{A5}$$

where $\mathcal{T}$ is the time-ordering operator. Similarly, the evolution operator from time τ to time t [the dashed part of the curve in Fig. 4(a)] is

$$\hat{U}(t;\tau) = \mathcal{T}\exp\left[-i\int_{\tau}^{t} d\tau' \hat{H}(\tau')\right]. \tag{A6}$$

If one now compares this to the evolution operator from time t_0 to time $\bar{\tau}$ in the system with the time-reversed Hamiltonian [the solid part of the curve in Fig. 4(b)],

$$\overline{\hat{U}(t;\bar{\tau})} = \mathcal{T}\exp\left[-i\int_{t_0}^{\bar{\tau}} d\tau' \,\overline{\hat{H}(\tau')}\right], \tag{A7}$$

where one should recall that $\bar{\tau} = t_0 + t - \tau$. Then it is straightforward to show that

$$\overline{\hat{U}(\bar{\tau};t_0)} = \hat{\Theta}^\dagger\, \hat{U}^\dagger(t;\tau)\, \hat{\Theta}. \tag{A8}$$

[1] See Chap. 6 of Gérard Battail, *Information and Life* (Springer, Dordrecht, 2014).

[2] B. Derrida, P. Gaspard, and C. Van den Broeck (eds.), Special issue: Work, dissipation, and fluctuations in nonequilibrium physics, C. R. Phys. **8**, 483 (2007).

[3] E. M. Sevick, R. Prabhakar, S. R. Williams, and D. J. Searles, Fluctuation theorems, Annu. Rev. Phys. Chem. **59**, 603 (2008).

[4] M. Campisi, P. Hänggi, and P. Talkner, Quantum fluctuation relations: Foundations and applications, Rev. Mod. Phys. **83**, 771 (2011).

[5] S. Vinjanampathy and J. Anders, Quantum thermodynamics, Contemp. Phys. **57**, 545 (2016).

[6] J. Millen and A. Xuereb, Perspective on quantum thermodynamics, New J. Phys. **18**, 011002 (2016).

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

[8] T. Schmiedl and U. Seifert, Stochastic thermodynamics of chemical reaction networks, J. Chem. Phys. **126**, 044101 (2007).

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

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

[11] G. Benenti, G. Casati, K. Saito, and R. S. Whitney, Fundamental aspects of steady-state conversion of heat to work at the nanoscale, Phys. Rep. **694**, 1 (2017).

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

[13] C. Jarzynski, Nonequilibrium Equality for Free Energy Differences, Phys. Rev. Lett. **78**, 2690 (1997); Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach, Phys. Rev. E **56**, 5018 (1997).

[14] D. J. Evans and D. J. Searles, Equilibrium microstates which generate second law violating steady states, Phys. Rev. E **50**, 1645 (1994).

[15] G. Crooks, Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences, Phys. Rev. E **60**, 2721 (1999).

[16] H. Tasaki, Jarzynski relations for quantum systems and some applications, arXiv:cond-mat/0009244.

[17] The notation $\Delta S_{\rm tot}$ for the total entropy change follows Refs. [7–9]; however other notations used include [10,46], $\Delta_i s$ or simply [11] $\Delta\mathcal{S}$.

[18] We take entropy in units of $k_{\rm B}$, temperature in units of energy, and time in units of 1/(energy); so $\hbar = k_{\rm B} = 1$.

[19] T. Yamada and K. Kawasaki, Nonlinear effects in the shear viscosity of critical mixtures, Prog. Theor. Phys. **38**, 1031 (1967).

[20] G. P. Morriss and D. J. Evans, Isothermal response theory, Mol. Phys. **54**, 629 (1985).

[21] D. M. Carberry, S. R. Williams, G. M. Wang, E. M. Sevick, and D. J. Evans, The Kawasaki identity and the fluctuation theorem, J. Chem. Phys. **121**, 8179 (2004).

[22] B. Roche, P. Roulleau, T. Jullien, Y. Jompol, I. Farrer, D. A. Ritchie, and D. C. Glattli, Harvesting dissipated energy with a mesoscopic ratchet, Nat. Commun. **6**, 6738 (2015).

[23] F. Hartmann, P. Pfeffer, S. Höfling, M. Kamp, and L. Worschech, Voltage Fluctuation to Current Converter with Coulomb-Coupled Quantum Dots, Phys. Rev. Lett. **114**, 146805 (2015).

[24] H. Thierschmann, R. Sánchez, B. Sothmann, F. Arnold, C. Heyn, W. Hansen, H. Buhmann, and L. W. Molenkamp, Three-terminal energy harvester with coupled quantum dots, Nat. Nanotechnol. **10**, 854 (2015).

[25] J.-L. Pichard and R. S. Whitney (eds.), Special issue: Mesoscopic thermoelectric phenomena, C. R. Phys. **17**, 1039 (2016).

[26] S. Juergens, F. Haupt, M. Moskalets, and J. Splettstoesser, Thermoelectric performance of a driven double quantum dot, Phys. Rev. B **87**, 245423 (2013).

[27] H. Schoeller and G. Schön, Mesoscopic quantum transport: Resonant tunneling in the presence of a strong Coulomb interaction, Phys. Rev. B **50**, 18436 (1994).

[28] J. König, J. Schmid, H. Schoeller, and G. Schön, Resonant tunneling through ultrasmall quantum dots: Zero-bias anomalies, magnetic-field dependence, and boson-assisted transport, Phys. Rev. B **54**, 16820 (1996).

[29] J. König, H. Schoeller, and G. Schön, Cotunneling at Resonance for the Single-Electron Transistor, Phys. Rev. Lett. **78**, 4482 (1997).

[30] H. Schoeller, Transport theory of interacting quantum dots, in *Mesoscopic Electron Transport*, edited by L. L. Sohn, L. P. Kouwenhoven, and G. Schön, NATO-ASI Series E: Applied Sciences, Vol. 345 (Springer, Dordrecht, 1997); online at http://digbib.ubka.uni-karlsruhe.de/volltexte/documents/2135.

[31] For a review, see R. Kosloff, Quantum thermodynamics: A dynamical viewpoint, Entropy **15**, 2100 (2013).

[32] G. Nenciu, Independent electron model for open quantum systems: Landauer-Büttiker formula and strict positivity of the entropy production, J. Math. Phys. **48**, 033302 (2007).

[33] R. S. Whitney, Thermodynamic and quantum bounds on nonlinear dc thermoelectric transport, Phys. Rev. B **87**, 115404 (2013).

[34] M. F. Ludovico, J. S. Lim, M. Moskalets, L. Arrachea, and D. Sánchez, Dynamical energy transfer in ac-driven quantum systems, Phys. Rev. B **89**, 161306(R) (2014).

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

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

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

[38] A. Kamenev, *Field Theory of Non-Equilibrium Systems* (Cambridge University Press, Cambridge, 2011).

[39] M. Leijnse and M. R. Wegewijs, Kinetic equations for transport through single-molecule transistors, Phys. Rev. B **78**, 235424 (2008).

[40] H. Schoeller, A perturbative nonequilibrium renormalization group method for dissipative quantum mechanics: Real-time RG in frequency space, Eur. Phys. J. Special Topics **168**, 179 (2009).

[41] R. B. Saptsov and M. R. Wegewijs, Time-dependent quantum transport: Causal superfermions, exact fermion-parity protected decay modes, and Pauli exclusion principle for mixed quantum states, Phys. Rev. B **90**, 045407 (2014).

[42] B. Sothmann, Electronic waiting-time distribution of a quantum-dot spin valve, Phys. Rev. B **90**, 155315 (2014).

[43] J. Schulenborg, R. B. Saptsov, F. Haupt, J. Splettstoesser, and M. R. Wegewijs, Fermion-parity duality and energy relaxation in interacting open systems, Phys. Rev. B **93**, 081411 (2016).

[44] E. B. Davies, Markovian master equations, Commun. Math. Phys. **39**, 91 (1974).

[45] E. B. Davies, Markovian master equations II, Math. Ann. **219**, 147 (1976).

[46] C. Elouard, D. Herrera-Mart, M. Clusel, and A. Aufféves, The role of quantum measurement in stochastic thermodynamics, NPJ Quantum Inf. **3**, 9 (2017).

[47] A. O. Caldeira and A. J. Leggett, Path integral approach to quantum Brownian motion, Phys. A (Amsterdam, Neth.) **121**, 587 (1983).

[48] A. O. Caldeira and A. J. Leggett, Quantum tunnelling in a dissipative system, Ann. Phys. **149**, 374 (1983).

[49] A. J. Leggett, S. Chakravarty, A. T. Dorsey, Matthew P. A. Fisher, Anupam Garg, and W. Zwerger, Dynamics of the dissipative two-state system, Rev. Mod. Phys. **59**, 1 (1987).

[50] See, e.g., the quotes from two letters from Maxwell to Tait (the first on 11 Dec 1867 and second undated), on pages 213-215 of C. G. Knott, *Life and Scientific Work of Peter Guthrie Tait* (Cambridge University Press, Cambridge, 1911).

[51] J. V. Koski, A. Kutvonen, I. M. Khaymovich, T. Ala-Nissila, and J. P. Pekola, On-Chip Maxwell's Demon as an Information-Powered Refrigerator, Phys. Rev. Lett. **115**, 260602 (2015).

[52] N. Cottet, S. Jezouin, L. Bretheau, P. Campagne-Ibarcq, Q. Ficheux, J. Anders, A. Auffèves, R. Azouit, P. Rouchon, and B. Huard, Observing a quantum Maxwell demon at work, Proc. Natl. Acad. Sci. USA **114**, 7561 (2017).

[53] Section 8.10 of the review Ref. [11].

[54] See, e.g., Chap. XV of A. Messiah, *Quantum Mechanics* (North Holland, Amsterdam, 1962).

[55] In fact, Crooks [15] did not consider classical systems without time-reversal symmetry (implicitly assuming no external magnetic fields), so he made no distinction between a state and its time reverse. However, it is a trivial step to note that, in general, the states on the left of his equality should be the time reverse of the states on the right.

[56] A. G. Redfield, On the theory of relaxation processes, IBM J. Res. Dev. **1**, 19 (1957).

[57] F. Bloch, Generalized theory of relaxation, Phys. Rev. **105**, 1206 (1957).

[58] S. Nakajima, On quantum theory of transport phenomena, Prog. Theor. Phys. **20**, 948 (1958).

[59] R. Zwanzig, Ensemble method in the theory of irreversibility, J. Chem. Phys. **33**, 1338 (1960).

[60] Chapter 15: Open quantum systems, in M. Le Bellac, *Quantum Physics* (Cambridge University Press, Cambridge, 2006).

[61] Chapter IV: Radiation considered as a reservoir: Master equation for the particles, in C. Cohen-Tannoudji, J. Dupont-Roc, and G. Grynberg, *Atom-Photon Interactions: Basic Process and Applications* (Wiley, New York, 1998).

[62] Chapter 8 of K. Blum, *Density Matrix Theory and Applications*, 3rd ed. (Springer-Verlag, Berlin, 2012). This was Chap. 7 in the first edition (1981).

[63] G. Lindblad, On the generators of quantum dynamical semigroups, Commun. Math. Phys. **48**, 119 (1976).

[64] R. S. Whitney, Staying positive: Going beyond Lindblad with perturbative master equations, J. Phys. A **41**, 175304 (2008).