Fully Quantum Fluctuation Theorems

Johan Åberg*
Institute for Theoretical Physics, University of Cologne, Zülpicherstrasse 77, D-500937, Cologne, Germany

Systems that are driven out of thermal equilibrium typically dissipate random quantities of energy on microscopic scales. Crooks fluctuation theorem relates the distribution of these random work costs with the corresponding distribution for the reverse process. By an analysis that explicitly incorporates the energy reservoir that donates the energy, and the control system that implements the dynamic, we here obtain a quantum generalization of Crooks theorem that not only includes the energy changes on the reservoir, but the full description of its evolution, including coherences. This approach moreover opens up for generalizations of the concept of fluctuation relations. Here we introduce 'conditional' fluctuation relations that are applicable to non-equilibrium systems, as well as approximate fluctuation relations that allow for the analysis of autonomous evolution generated by global time-independent Hamiltonians.

arXiv:1601.01302v1 [quant-ph] 6 Jan 2016

I. INTRODUCTION

Imagine a physical system with a Hamiltonian $H_S(x)$ that depends on some external parameter x, e.g., electric or magnetic fields that we can vary at will. By changing x we can push the system out of thermal equilibrium. This would typically require work that may be dissipated due to interactions with the surrounding heat bath. The latter may also make the dissipation random, in the sense that the work cost w is different each time we implement the same change of the Hamiltonian [1–6]. (Think of a spoon pushed through syrup. On microscopic scales the friction resolves into random molecular collisions.) These fluctuations can be described via a probability distribution $P_+(w)$. Crooks theorem [7] shows that there is a surprisingly simple relation between P_+ and the corresponding distribution P_- obtained for the process run in reverse (i.e., where the time-schedule for the change of x is mirrored) namely

$$Z(H_S^i)P_+(w) = Z(H_S^f)e^{\beta w}P_-(-w), \tag{1}$$

where H_S^i and H_S^f are the initial and final Hamiltonians, respectively, and $Z(H)$ is the partition function, which in the quantum case takes the form $Z(H) = \mathrm{Tr}e^{-\beta H}$. Moreover, $\beta = 1/(kT)$, with k Boltzmann's constant, and where T is the absolute temperature of the heat bath. In this investigation we only consider a single heat bath with a fixed temperature T.

Crooks theorem was originally derived in a classical setting [7] (for reviews on classical fluctuation theorems, see [1–6]) but there has accumulated a considerable body of quantum fluctuation relations (for reviews see [8–10]). However, the latter do often include measurements, and in particular energy measurements (see e.g. [11–25]) that typically destroy coherences and quantum correlations. Here we avoid such auxiliary components, and obtain fluctuation relations that retain all quantum aspects of the evolution.

The key is to explicitly model all degrees of freedom involved in the process. This includes the control mechanism that implements the change of the external parameter, i.e. x in $H_S(x)$, as well as the 'energy reservoir', e.g., a battery or an excited atom, which donates the energy for the work w. Since work corresponds to a change of energy in this reservoir, it follows that the reservoir itself is forced to evolve as it fuels the dynamics. The general theme of this investigation is to formulate fluctuation relations in terms of this induced evolution. To make this more concrete, let us briefly display the first of the fluctuation theorems that we will derive. The roles of the probability distributions P_+ and P_- are here taken over by the channels $\mathcal{F}_+$ and $\mathcal{F}_-$ that are induced on the energy reservoir by the forward and reverse processes, respectively. As we shall see, these channels can, under suitable conditions, be related by the following quantum Crooks relation

$$Z(H_S^i)\mathcal{F}_+ = Z(H_S^f)\mathcal{J}_{\beta H_E}\mathcal{F}_-^{\ominus}\mathcal{J}_{\beta H_E}^{-1}, \tag{2}$$

where H_E denotes the Hamiltonian of the energy reservoir E. The combined application of $\mathcal{J}_{\beta H_E}$ and $\mathcal{J}_{\beta H_E}^{-1}$, where $\mathcal{J}_{\beta H_E}(Q) = e^{-\beta H_E/2}Qe^{-\beta H_E/2}$, can be viewed as a counterpart to the term $e^{\beta w}$ in (1). The mapping from $\mathcal{F}_-$ to $\mathcal{F}_-^{\ominus}$ (to be described in detail later) is related to time-reversals, and can in some sense be regarded as a generalization of the transformation of $P_-(w)$ to $P_-(-w)$ at the right hand side of (1).

In the following section we derive (2) and moreover show that it can be decomposed into diagonal and off-diagonal Crooks relations, thus yielding fluctuation relations for coherences. We also derive quantum Jarzynski equalities, as well as a bound on the average work cost. In section II H we regain the classical Crooks relation (1) from (2) via the additional assumption of energy translation invariance. We next turn to generalizations of (2), where section III introduces conditional fluctuation relations, and section VI approximate versions.

*Electronic address: johan.aberg@uni-koeln.de

II. A QUANTUM FLUCTUATION THEOREM

A. The model

To derive the quantum Crooks relation in (2) we employ a general class of models that previously has been used in the context of quantum thermodynamics to analyze, e.g., work extraction, information erasure, and coherence [26–43]. The main idea is that we include all the relevant degrees of freedom (which in our case consist of four subsystems, see Fig. 1) and assign a global time-independent Hamiltonian H to account for energy. On this joint system we are allowed to act with any unitary operation V that conserves energy, which is formalized by the condition that V commutes with the total Hamiltonian $[H, V] = 0$. (See [44] for an alternative notion of energy conservation.)

At first sight it may be difficult to see how such a manifestly time-independent Hamiltonian can be used to describe the evolving Hamiltonians in Crooks theorem. For this purpose we introduce a control system C such that the Hamiltonian of $S' = SB$ depends on the state of C. As an illustration, suppose that we wish to describe a transition from an initial Hamiltonian $H^i_{S'}$ to a final Hamiltonian $H^f_{S'}$. One possibility would be to define a joint Hamiltonian of the form $H_{S'C} = H^f_{S'} \otimes |c_f\rangle\langle c_f| + H^i_{S'} \otimes |c_i\rangle\langle c_i|$, where $|c_f\rangle$ and $|c_i\rangle$ are two normalized and orthogonal states of the control system C. If the evolution would change the control from $|c_i\rangle$ to $|c_f\rangle$, then this would effectively change the Hamiltonian of S' from $H^i_{S'}$ to $H^f_{S'}$. However, this change can in general not be achieved by an energy conserving unitary operation on $S'C$ alone, since the transition from $H^i_{S'}$ to $H^f_{S'}$ typically will involve a change of energy. The role of the energy reservoir E is to make these transitions possible by donating or absorbing the necessary energy. With a suitable choice of Hamiltonian H_E of the energy reservoir, the global Hamiltonian $H = H_{S'C} \otimes \hat{1}_E + \hat{1}_{S'C} \otimes H_E$ allows for non-trivial energy conserving unitary operators V.

In this investigation the energy reservoir does not only serve as a source or sink for energy, but also acts as a probe of the dynamics of the system. Instead of using auxiliary measurements, which inevitably would introduce additional interactions that potentially may disturb the dynamics, we here let the energy reservoir take over this role. We do, so to speak, make an additional use of a component that anyway has to be included. A further benefit of using an explicit quantum probe is that we not only capture the flow of energy, but also the change in coherence and correlations. Needless to say, standard macroscopic energy reservoirs would be of little use for the latter purposes, due to the levels of decoherence that they are exposed to. However, our notion of energy reservoirs includes nano-systems like, e.g., single atoms or spins.

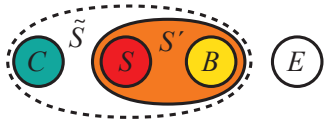

FIG. 1: **The systems.** The model explicitly includes the 'system' S on which we operate, the heat bath B, the control C that implements the change of the Hamiltonian on SB, and the energy reservoir E that donates or accepts the energy required to drive the processes. Most of our fluctuation relations are expressed in terms of the dynamics induced on the energy reservoir E.
For the main part of this investigation (Secs. II and III) we assume that the total system is described by a time-independent Hamiltonian that is non-interacting between SBC and E, i.e., is of the form $H = H_{SBC} \otimes \hat{1}_E + \hat{1}_{SBC} \otimes H_E$. We moreover model the evolution by energy conserving unitary operators V, i.e., $[H, V] = 0$. (We go beyond these assumptions when we consider approximate fluctuation relations in section VI.) For most of the derivations it is not necessary to make any distinction between S and B, and for this reason we will often bundle them together into an extended system $S' = SB$. This 'rationalization' can be taken one step further to $\tilde{S} = S'C = SBC$ when we turn to the conditional fluctuation theorems in section III.

B. An intermediate version

In this section we derive a simplified version of our fluctuation theorem. This serves as a convenient intermediate step, and illustrates why time-reversal symmetry is useful.

We wish to determine the channel induced on the energy reservoir by a non-equilibrium process *a la* Crooks, where the system initially is in equilibrium, and is forced out of it. To model this we let the control system initially be in state $|c_i\rangle$, and we let the combined system and heat bath S' be in the Gibbs state $G(H^i_{S'})$, where $G(H) = e^{-\beta H}/Z(H)$. Moreover, we let the reservoir E be in an arbitrary state σ. Hence, the joint system is initially in the state $G(H^i_{S'}) \otimes |c_i\rangle\langle c_i| \otimes \sigma$. After a global energy-conserving unitary operation V, the state of the energy reservoir is consequently given by the channel

$$\mathcal{F}(\sigma) = \mathrm{Tr}_{S'C}\big(V[G(H^i_{S'}) \otimes |c_i\rangle\langle c_i| \otimes \sigma]V^\dagger\big). \quad (3)$$

To obtain a Crooks-like relation we should somehow relate this forward process with a 'reversed' process, where the system instead is initiated in the Gibbs state of the final Hamiltonian, i.e., the global system should be in the joint state $G(H^f_{S'}) \otimes |c_f\rangle\langle c_f| \otimes \sigma$. The question is how to formalize the idea of a reversed process acting on this initial state. A rather brutal interpretation would be to simply invert the entire global evolution, thus substituting V with $V^\dagger$ (see Fig. 2). The resulting channel on the reservoir would in this case be

$$\mathcal{R}(\sigma) = \mathrm{Tr}_{S'C}\big(V^\dagger[G(H^f_{S'}) \otimes |c_f\rangle\langle c_f| \otimes \sigma]V\big). \quad (4)$$

Let us now assume that we have a perfectly functioning control system, in the sense that the state $|c_i\rangle$ is

transformed into $|c_f\rangle$ with certainty. More precisely, we demand that the unitary operator V should satisfy the condition

$$[\hat{1}_{S'} \otimes |c_f\rangle\langle c_f| \otimes \hat{1}_E]V = V[\hat{1}_{S'} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E]. \quad (5)$$

One can verify (see Appendix A) that this assumption implies the following relation between the channels $\mathcal{F}$ and $\mathcal{R}$

$$Z(H_S^i)\mathcal{F} = Z(H_S^f)\mathcal{J}_{\beta H_E}\mathcal{R}^*\mathcal{J}_{\beta H_E}^{-1}. \quad (6)$$

Here $\mathcal{R}^*$ denotes the conjugate [45] of the channel $\mathcal{R}$. If the channel has Kraus representation $\mathcal{R}(\rho) = \sum_j V_j \rho V_j^\dagger$, then the conjugate can be written $\mathcal{R}^*(\rho) = \sum_j V_j^\dagger \rho V_j$ [45].

As the reader may have noticed, we have written $Z(H_S^i)$ and $Z(H_S^f)$ in (6) rather than the more generally valid $Z(H_{S'}^i)$ and $Z(H_{S'}^f)$. To obtain the former we can assume that $H_{S'}^i = H_S^i \otimes \hat{1}_B + \hat{1}_S \otimes H_B$ and $H_{SB}^f = H_S^f \otimes \hat{1}_B + \hat{1}_S \otimes H_B$, and thus $Z(H_{S'}^i) = Z(H_S^i)Z(H_B)$ and $Z(H_{S'}^f) = Z(H_S^f)Z(H_B)$.

One may observe that the right hand side of (6) is similar to Crooks' quantum operation time reversal [46] (not to be confused with the time reversals $\mathcal{T}$ or the mapping $\ominus$ discussed in the next section) and closely related to Petz recovery channel [47–50], where a relation to work extraction was identified recently [40]. For further comments, see Appendix A 8.

The approach of this investigation can be compared with another construction that also focuses on quantum channels [51, 52], where the starting point is to assume a property of a channel (unitality) and derive fluctuation relations for the resulting probability distribution for suitable classes of initial and final measurements that sandwich the channel (see also the generalization in [53]).

Although (6) indeed can be regarded as a kind of quantum Crooks relation, it does suffer from an inherent flaw. The swap between V and $V^\dagger$ means that we invert the entire evolution of all involved systems, including the heat bath. Apart from assuming an immense level of control, this assumption does not quite fit with the spirit of Crooks relation. The latter only assumes a reversal of the time-schedule of the control parameters, not a reversal of the entire evolution. In the following we will resolve this issue by invoking time-reversal symmetry.

C. Time-reversals and time-reversal symmetry

In standard textbooks on quantum mechanics (see e.g. [54]) time-reversals are often introduced on the level of Hilbert spaces via complex conjugation of wavefunctions (also applied in the context of quantum fluctuation relations, see e.g. [9, 14, 22, 24, 55]). Here we instead regard time-reversals as acting on operators (cf. [56, 57]). Moreover, our time-reversals have the flavor of transpose

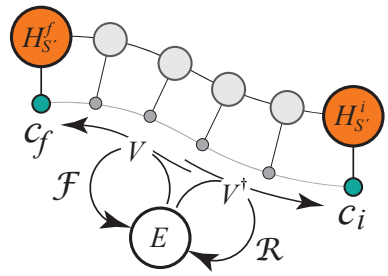

FIG. 2: **An intermediate quantum Crooks relation.** Our quantum Crooks theorem is based on the idealized idea of a perfect control mechanism. This assumes that the energy conserving unitary evolution V on the joint system $S'CE$ turns the control state $|c_i\rangle$ into $|c_f\rangle$ with certainty, thus implementing the change of the Hamiltonian $H_{S'}^i$ to $H_{S'}^f$ perfectly. To model the forward process of the Crooks relation we assume that $S'C$ starts in the 'conditional' equilibrium state $G(H_{S'}^i) \otimes |c_i\rangle\langle c_i|$. The subsequent evolution under V induces a channel $\mathcal{F}$ on the energy reservoir E. As an intermediate step towards the quantum Crooks relation (2) we assume that the reverse process is given by the globally reversed evolution $V^\dagger$. With $S'C$ in the initial state $G(H_{S'}^f) \otimes |c_f\rangle\langle c_f|$, this results in a channel $\mathcal{R}$ on E. The preliminary Crooks relation in Eq. (6) relates the channels $\mathcal{F}$ and $\mathcal{R}$.
This result does not rely on any additional assumptions on how V comes about, or on the nature of the dynamics at intermediate times. However, if one so wishes, it is possible to explicitly model the path $H(x)$ in the space of Hamiltonians. This path can be discretized as H_l for $l = 0, \ldots, L$ with $H_0 = H_{S'}^i$ and $H_L = H_{S'}^f$. To these Hamiltonians we associate states $|c_l\rangle$ in the control C, and construct the joint Hamiltonian $H_{S'C} = \sum_{l=0}^L H_l \otimes |c_l\rangle\langle c_l|$. Hence, if the evolution traverses the family of control states, then we progress along the path of Hamiltonians. An alternative method to model evolving Hamiltonians is discussed in section VI.

operations rather than complex conjugations. Since density operators and observables are Hermitian, the choice between conjugation or transposition is largely a matter of taste. Here we opt for the transpose, since this choice avoids the inconvenient anti-linearity of complex conjugation. A couple of key properties of our time-reversals $\mathcal{T}$ are $\mathcal{T}(AB) = \mathcal{T}(B)\mathcal{T}(A)$ and $\mathcal{T}(A)^\dagger = \mathcal{T}(A^\dagger)$. (For the complete characterization, see Appendix B.)

To see why it is reasonable to refer to $\mathcal{T}$ as a 'time-reversal', let us assume that the evolution operator V is time-reversal invariant, i.e., $\mathcal{T}(V) = V$. If an initial state ρ_i is evolved into $\rho_f = V\rho_i V^\dagger$, then the properties of $\mathcal{T}$ mentioned above yield $\mathcal{T}(\rho_i) = V\mathcal{T}(\rho_f)V^\dagger$. In other words, the reversed final state state $\mathcal{T}(\rho_f)$ evolves to the reversed initial state $\mathcal{T}(\rho_i)$. Note that this 'backwards' transformation is implemented by the forward evolution $V\rho V^\dagger$, rather than the inverted evolution $V^\dagger \rho V$.

The $\ominus$-transformation that appears in (2) can be defined via the chosen time-reversal as

$$\mathcal{F}^\ominus = \mathcal{T}\mathcal{F}^*\mathcal{T}. \quad (7)$$

As mentioned earlier, if $\mathcal{F}(\sigma) = \sum_j V_j \sigma V_j^\dagger$, then the con-

jugate is given by $\mathcal{F}^*(\sigma) = \sum_j V_j^\dagger \sigma V_j$. By the properties of $\mathcal{T}$ mentioned above it thus follow that $\mathcal{F}^\ominus(\sigma) = \sum_j \mathcal{T}(V_j)\sigma\mathcal{T}(V_j)^\dagger$, i.e., we time-reverse the operators in the Kraus representation.

In the finite-dimensional case each time-reversal can be implemented by a transpose with respect to some orthonormal basis, followed by a special class of unitary operations. (Hence, our time-reversals do strictly speaking include a bit larger set of operations than proper transposes, see Appendix B.) Since the transpose is a positive but not completely positive map (as is illustrated by the effect of the partial transpose on entangled states [58, 59]) we can conclude that time-reversals generally are not physical operations. Hence, one should not be tempted to think of time-reversals as something that we actually apply to a system. However, given a *description* of a state ρ, there is nothing that in principle prevents us from preparing the time reversed state $\mathcal{T}(\rho)$, which is sufficient for our purposes.

D. Deriving the quantum Crooks relation

Similar to our intermediate version in section II B we here assume a non-interacting global Hamiltonian $H = H_{S'C}\otimes\hat{1}_E + \hat{1}_{S'C}\otimes H_E$, and a global energy conserving unitary evolution $[V, H] = 0$. As a substitute for the global inversion of the evolution we assume time-reversal symmetry in terms of a suitably chosen time-reversal of the the form $\mathcal{T} = \mathcal{T}_{S'C} \otimes \mathcal{T}_E$, i.e., the global time reversal is composed of local reversals on $S'C$ and on E. We impose the time reversal symmetry by assuming that $\mathcal{T}(V) = V$, $\mathcal{T}_{S'C}(H_{S'C}) = H_{S'C}$, and $\mathcal{T}_E(H_E) = H_E$.

Another novel component compared to the intermediate version is that we associate *pairs* of control states to the Hamiltonians on S'. We assign the pair $|c_{i+}\rangle$, $|c_{i-}\rangle$ for the initial Hamiltonian $H^i_{S'}$, and the pair $|c_{f+}\rangle$, $|c_{f-}\rangle$ for the final Hamiltonian $H^f_{S'}$. The general idea is that the members of these pairs correspond to the forward and reverse evolution of the control system. One way to formalize this notion is via the following relations

$$\begin{aligned}\mathcal{T}_{S'C}(\hat{1}_{S'} \otimes |c_{i+}\rangle\langle c_{i+}|) &= \hat{1}_{S'} \otimes |c_{i-}\rangle\langle c_{i-}|,\\ \mathcal{T}_{S'C}(\hat{1}_{S'} \otimes |c_{f+}\rangle\langle c_{f+}|) &= \hat{1}_{S'} \otimes |c_{f-}\rangle\langle c_{f-}|.\end{aligned} \tag{8}$$

Hence, the time-reversal swaps the two control states into each other. Note that if V transforms $|c_{i+}\rangle$ to $|c_{f+}\rangle$ perfectly, then V also transforms $|c_{f-}\rangle$ to $|c_{i-}\rangle$ perfectly. In the latter case one should keep in mind the reverse ordering of i and f; the state $|c_{f-}\rangle$ is the initial control state of the reverse process (see Fig. 3).

Let us assume that the unitary operator V and the initial and final control states $|c_{i+}\rangle, |c_{f+}\rangle$ satisfy the conditions of the intermediate fluctuation relation. The channels $\mathcal{F}_+$ and $\mathcal{R}_+$, obtained by substituting $|c_i\rangle$ with $|c_{i+}\rangle$, and $|c_f\rangle$ with $|c_{f+}\rangle$ in equations (3) and (4), respectively, are according to (6) related by

$$Z(H^i_{S'})\mathcal{F}_+ = Z(H^f_{S'})\mathcal{J}_{\beta H_E}\mathcal{R}^*_+\mathcal{J}^{-1}_{\beta H_E}. \tag{9}$$

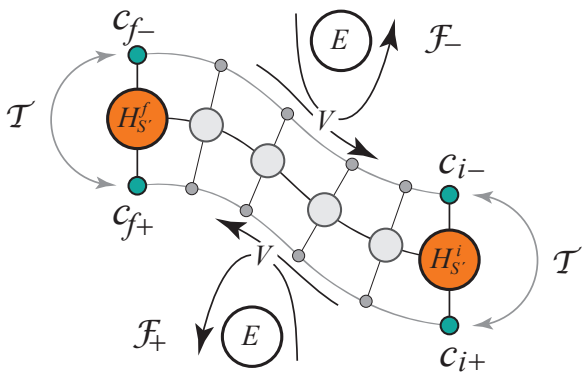

FIG. 3: **A quantum Crooks relation.** Derivations of fluctuation relations often rely on time-reversals and time reversal symmetry in some form [3–6, 8–10]. In our case, time-reversal symmetry enables us to in effect let the evolution run 'backwards' by swapping the control states from a 'forward track' to a 'reverse track'. By this we amend for the over-ambitious global reversal of the evolution that we employed for the intermediate fluctuation relation, and obtain a quantum counterpart to the reversal of the control parameter in the classical Crooks relation.
Analogous to the setup in Fig. 2, the initial control state $|c_{i+}\rangle$ yields an evolution where the initial Hamiltonian $H^i_{S'}$ is transformed to the final Hamiltonian $H^f_{S'}$ when the control state $|c_{f+}\rangle$ is reached. This process induces a channel $\mathcal{F}_+$ on the energy reservoir E. To each of these control states there exists a time-reversed partner starting with $|c_{f-}\rangle$ and ending with $|c_{i-}\rangle$ thus bringing the Hamiltonian $H^f_{S'}$ back to $H^i_{S'}$. This process induces a channel $\mathcal{F}_-$ on the reservoir. By assuming time-reversal symmetry, the channels $\mathcal{F}_+$ and $\mathcal{F}_-$ can be related via the quantum Crooks relation in (2).

For the reverse process we obtain the channels $\mathcal{F}_-$ and $\mathcal{R}_-$, where $|c_i\rangle$ this time is substituted by $|c_{f-}\rangle$, and $|c_f\rangle$ with $|c_{i-}\rangle$. By using the the properties of the time-reversal one can show (for details, see Appendix C) that $\mathcal{R}_+\mathcal{T}_E = \mathcal{T}_E\mathcal{F}_-$. By applying this equality to (9) our desired quantum Crooks relation (2) follows immediately. Note that both $\mathcal{F}_+$ and $\mathcal{F}_-$ correspond to the 'forward' evolution, i.e. we no longer need to impose a global inversion of the dynamics.

After having established (2) an immediate question is how to interpret it physically. The classical Crooks relation (1) can be phrased as a comparison between two experiments, one for the forward path of control parameters, and one for the reversed. For each of these two setups we repeat the experiment in order to obtain good statistical estimates of the probability distributions P_+ and P_-. The claim, so to speak, of Crooks theorem is that these two probability distributions are related as in (1). Our quantum fluctuation relation in (2) can be interpreted in a similar manner. Here the channels $\mathcal{F}_+$ and $\mathcal{F}_-$ induced by the forward and reverse process can in principle be determined by process tomography (see e.g. [60–63] and Sec. 8.4.2 in [64] for an overview). Analogous to the classical Crooks relation, our quantum fluc-

tuation theorem tells us that these two channels should be related as in (2). One may note that the mapping $\mathcal{F}_- \mapsto \mathcal{J}_{\beta H_E}\mathcal{F}_-^{\ominus}\mathcal{J}_{\beta H_E}^{-1}$ in this setup is something that we *calculate* given an experimental estimate of the channel $\mathcal{F}_-$, analogous to how we would calculate the function $e^{\beta w}P_-(-w)$ from the estimated distribution $P_-(w)$. In section IV we consider a reformulation of (2) that is associated with a scenario that does not require such a post-processing.

As the reader may have noticed, our quantum Crooks relation (2) does, in contrast to the classical Crooks relation (1), not contain any explicit reference to 'work'. Hence, we do in essence circumvent the issue of how to translate the classical notion of work into the quantum setting; a question that has been discussed rather extensively in relation to fluctuation theorems [11, 12, 15, 65–78]. We can nevertheless in some sense associate work with the loss of energy in the reservoir, which we employ in section II H where we regain the classical Crooks relation (1).

E. Diagonal and off-diagonal Crooks relations

Here we take a closer look at the nature of the induced channels $\mathcal{F}_\pm$ and on the structure of the quantum Crooks relation (2).

The energy conservation and the diagonal initial state with respect to the Hamiltonian of SBC conspire to decouple the dynamics on E with respect to the 'modes of coherence' [34] (which in turn is a part of the wider context of symmetry preserving operations [79]). To make this more concrete, assume that H_E has a complete set of energy eigenstates $|n\rangle$ with corresponding non-degenerate energies E_n. Let us also assume that $\mathcal{T}_E$ acts as the transposition with respect to this energy eigenbasis. The decoupling does in this case mean that the channels $\mathcal{F}_\pm$ can map an element $|n\rangle\langle n'|$ to $|m\rangle\langle m'|$ only if $E_n - E_{n'} = E_m - E_{m'}$ (see Appendix D for further details). Hence, each mode of coherence is characterized by an 'offset' $\delta = E_n - E_{n'}$ from the main diagonal (see Fig. 4).

In particular, the main diagonal is given by the offset $\delta = 0$. The dynamics of the diagonal elements under the forward process can be described via a conditional probability distribution

$$p_\pm(m|n) = \langle m|\mathcal{F}_\pm(|n\rangle\langle n|)|m\rangle = \langle n|\mathcal{F}_\pm^*(|m\rangle\langle m|)|n\rangle. \tag{10}$$

The application of (2) yields

$$Z(H_S^i)p_+(m|n) = e^{\beta(E_n - E_m)}Z(H_S^f)p_-(n|m), \tag{11}$$

which thus can be regarded as a diagonal Crooks relation. In section II H we shall see how this in turn can be used to re-derive (1) via the additional assumption of energy translation invariance.

For each permissible value of δ (which depend on the spectrum of H_E) we can characterize the dynamics of the

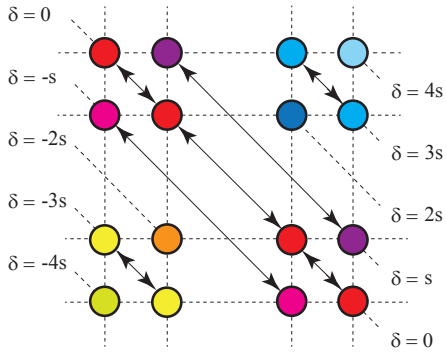

FIG. 4: **Diagonal and off-diagonal Crooks relations.** Due to energy conservation in combination with the particular class of initial states, it turns out that the dynamics of the energy reservoir decouples along the modes of coherence [34]. One can think of these as as the collection of diagonals (main and off-diagonals) of the density operator represented in an energy eigenbasis. Due to the decoupling it follows that the quantum fluctuation relation Eq. (2) can be separated into individual Crooks relations for each mode of coherence. As an example, suppose that the spectrum of the reservoir would include the energy levels $E_1 = 0, E_2 = s, E_3 = 3s, E_4 = 4s$ for some $s > 0$. The dynamics on the main diagonal, $\delta = 0$, satisfies the relation in (11). Each of the diagonals with offsets $\delta = \pm s, \pm 2s, \pm 3s, \pm 4s$ satisfies a Crooks relation as in (12).

corresponding off-diagonal mode by

$$q_\pm^\delta(m|n) = \langle m|\mathcal{F}_\pm(|n\rangle\langle n'|)|m'\rangle = \langle n'|\mathcal{F}_\pm^*(|m'\rangle\langle m|)|n\rangle,$$

where the assumption of non-degeneracy implies that n' is uniquely determined by n, δ, and similarly m' by m, δ. We obtain an off-diagonal analogue of (11) in the form of

$$Z(H_S^i)q_+^\delta(m|n) = e^{\beta(E_n - E_m)}Z(H_S^f)q_-^\delta(n|m). \tag{12}$$

In other words, the functional form of the relations for the main diagonal and all the off-diagonals are identical. However, one should keep in mind that the numbers $q_\pm^\delta(m|n)$ in general are complex, and thus cannot be interpreted as conditional probability distributions.

The conditional distributions $p_\pm(m|n)$ and all $q_\pm^\delta(m|n)$ can be regarded as representing different aspects of the channels $\mathcal{F}_\pm$. However, one can also give more direct physical interpretations to these quantities. To this end one should keep in mind that the energy reservoir can be viewed as being part of our experimental equipment, and we thus in principle are free to prepare and measure the reservoir it in any way that we wish. The conditional distribution $p_\pm(m|n)$ can be interpreted as the probability to measure the energy E_m in the reservoir after the the experiment has been completed, given that we prepared the reservoir in energy E_n before we connected it to the system. Equivalently, we could replace the initial preparation with an energy measurement, thus approaching the 'two-time' or 'two-point' measurements that play a central role in several investigations on quantum fluctuation relations (see e.g. [11–25], and for overviews see

[8, 9]), although we here apply these measurements to the energy reservoir, rather than on the system. An obvious issue with sequential measurements is that the first measurement typically perturbs the statistics of the second measurement. However, in our case the decoupling comes to our aid. If we measure the energy in the reservoir after the process, then the statistics of this measurement is not affected if we would insert an additional energy measurement before the process, irrespective of how off-diagonal the initial state may be. The decoupling thus justifies the use of a two-point measurement on the energy reservoir, if our purpose is to determine $p_\pm$. However, the very same argument also implies that if we wish to determine $q_\pm^\delta$ we have to use non-diagonal initial states, as well as non-diagonal measurements. For an example of such a setup, see Appendix D 3.

F. Jarzynski equalities

From the classical Crooks relation one can directly derive the Jarzynski equality $\langle e^{-\beta W}\rangle = Z(H^f)/Z(H^i)$ [6, 80]. As mentioned above, it is not clear how to translate the random variable W into the quantum setting, and consequently it is also not evident how to translate the expectation value $\langle e^{-\beta W}\rangle$ (which yields a variety of approaches to the quantum Jarzynski equality, see overviews in [8–10]). Like for the Crooks relation, we do not attempt a direct translation, but rather focus on the general dynamics in the energy reservoir.

By applying (2) to an initial state σ of the reservoir and take the expectation value of the operator $e^{\beta H_E}$, it follows that

$$\frac{\mathrm{Tr}[e^{\beta H}\mathcal{F}_+(\sigma)]}{\mathrm{Tr}[e^{\beta H_E/2}\mathcal{R}_+(\hat{1})e^{\beta H_E/2}\sigma]} = \frac{Z(H_S^f)}{Z(H_S^i)}. \tag{13}$$

Although this equality has the flavor of a Jarzynski equality, we can bring it one step further by additionally assuming that $\mathcal{R}_+(\hat{1}) = \hat{1}$, or equivalently $\mathcal{F}_-(\hat{1}) = \hat{1}$, i.e., that these channels are unital. This assumption yields

$$\frac{\mathrm{Tr}[e^{\beta H_E}\mathcal{F}_+(\sigma)]}{\mathrm{Tr}(e^{\beta H_E}\sigma)} = \frac{Z(H_{S'}^f)}{Z(H_{S'}^i)}. \tag{14}$$

This relation has a clear physical interpretation in terms of an experiment where we measure the expectation value of $e^{\beta H_E}$ on the input state of the reservoir (e.g, from the statistics of energy measurements), and in a separate experiment measure the same observable on the evolved state $\mathcal{F}_+(\sigma)$.

In a similar fashion one can derive a whole family of Jarzynski-like equalities (with as well as without the assumption $\mathcal{R}_+(\hat{1}) = \hat{1}$, see Appendix E) and another member of this family is

$$\mathrm{Tr}[e^{\beta H_E}\mathcal{F}_+(e^{-\beta H_E/2}\sigma e^{-\beta H_E/2})] = \frac{Z(H_S^f)}{Z(H_S^i)}. \tag{15}$$

This equality does not have a quite as direct physical interpretation as (14), but one can resort two-point energy measurements, resulting in fluctuation relations akin to e.g. [11–13, 76, 77]. To see this, assume that H_E has a point spectrum, then equation (15) takes the form $\sum_{mn} e^{\beta(E_m-E_n)}p_+(m|n)\langle n|\sigma|n\rangle = Z(H_S^f)/Z(H_S^i)$, where we have made use of the decoupling of the diagonal mode of coherence. This can alternatively be obtained directly from the diagonal Crooks relation (11), where the unitality of $\mathcal{F}_-$ implies that p_- is not only stochastic, but doubly stochastic.

G. Bound on the work cost

As mentioned above, the classical Jarzynski equality follows directly from Crooks relation. It is therefore perhaps a bit surprising that the condition of unitality of the channel $\mathcal{R}_+$ emerges in the derivations of our Jarzynski equalities. To explore this further, let us consider the expected work cost of these processes. For a process that starts in equilibrium we would expect (e.g as a special case of more general bounds on work costs [81–84]) that the average work cost should satisfy the bound $\langle W\rangle \geq F(H^f) - F(H^i)$, i.e., the work cost is at least equal to the difference in (equilibrium) free energy between the final and initial Hamiltonian (where equality typically is reached in the limit of quasi-static processes). In the classical case, this bound follows from Jarzynski's equality by using the convexity of the exponential function [80].

In our setting one can use the diagonal Crooks relation (11) to show that

$$\begin{aligned}\mathrm{Tr}(H_E\sigma) - \mathrm{Tr}\big(H_E\mathcal{F}_+(\sigma)\big) \geq & F(H_{S'}^f) - F(H_{S'}^i) \\ & - kT\ln\mathrm{Tr}\big(\sigma\mathcal{R}_+(\hat{1}_E)\big).\end{aligned} \tag{16}$$

(As a technical remark, we do for this derivation additionally assume that H_E has a pure non-degenerate point spectrum. For details see Appendix F.) If we identify the decrease in average energy of the reservoir, $\mathrm{Tr}(H_E\sigma) - \mathrm{Tr}\big(H_E\mathcal{F}_+(\sigma)\big)$ with the average work cost $\langle W\rangle$, we thus regain the standard bound if $\mathcal{R}_+$ is unital. Hence, the unitality of $\mathcal{R}_+$ appear again, this time to guarantee the standard work bound.

The inequality (16) does not necessarily mean that the standard bound is violated. However, one can construct an explicit example where the work cost is smaller than the standard bound for a process with a non-unital $\mathcal{R}_+$ (see Appendix F 2). At first this may seem a bit alarming since it would appear to suggest violations of basic thermodynamics. However, one has to keep in mind that we here include the energy reservoir as a physical system, and that we cannot expect to regain standard bounds if we allow ourselves to use the energy reservoir *per se* as a resource (cf. discussions in [44]), and one may suspect that the unitality of $\mathcal{R}_+$ is related to this issue. A further

indication in this direction is that the energy translation invariance, which is the topic of the next section, not only allows us to re-derive the classical Crooks relation (1), but also guarantees that the channel $\mathcal{R}_+$ is unital (see Appendix G).

H. Energy translation invariance: A bridge to standard fluctuation relations

Let us for a moment reconsider the classical Crooks relation in equation (1). As formulated, it makes no explicit reference to any energy reservoir. Moreover, the distributions $P_\pm$ are typically phrased as functions only of the work w, and thus only functions of the *change* of energy in the implicit reservoir (although it is difficult to see that anything in principle would prevent $P_\pm$ from depending on additional parameters, like the actual energy level of the reservoir). One could even argue that for a well designed experiment $P_\pm$ should not depend on how much energy there is in the reservoir (as long as there is enough), since this in some sense is a property of the experimental equipment rather than a property of the system under study. This is in contrast with our more general formulation in terms of channels on the reservoir, which very explicitly allows for the possibility that the exact choice of initial state of the reservoir (and thus the energy) can make a difference. In order to formalize the idea that the experiment should be independent of the amount of energy in the reservoir, we do in this section assume energy translation invariance. This enables us to derive the classical Crooks relation (1) from the quantum relation (2).

The technique for implementing energy translation invariance has previously been used to study work extraction and coherence [31, 44]. Here we use a model where the spectrum of H_E forms an equi-spaced doubly infinite energy ladder [31] (see also the continuum version in [44]), i.e., the energy eigenstates $|n\rangle$ correspond to energies $E_n = sn$ for some fixed number $s > 0$ and $n \in \mathbb{Z}$. We impose the energy translation invariance by assuming that the unitary operators V should commute with energy translations along the energy-ladder, $[V, \Delta] = 0$, where $\Delta = \sum_n |n+1\rangle\langle n|$. (Technically, we also assume that the eigenvalues of $H^i_{S'}$ and $H^f_{S'}$ are multiples of s.) As a consequence, the channels $\mathcal{F}_\pm$ also become energy-translation invariant in the sense that

$$\Delta^j \mathcal{F}_\pm(\sigma) {\Delta^\dagger}^k = \mathcal{F}_\pm\big(\Delta^j \sigma {\Delta^\dagger}^k\big), \tag{17}$$

for all $j, k \in \mathbb{Z}$ (see Appendix G). A consequence is that the diagonal transition probabilities $p_\pm$, defined in (10), inherit the translation invariance

$$p_\pm(m|n) = p_\pm(m-n|0) = p_\pm(0|n-m). \tag{18}$$

Moreover, the translation invariance (17) conspires with the decoupling of the diagonals decribed in section II E such that we can rewrite $\mathcal{F}_\pm$ as

$$\mathcal{F}_\pm(\rho) = \sum_{n,n',k\in\mathbb{Z}} p_\pm(k|0)\langle n|\rho|n'\rangle |n+k\rangle\langle n'+k|. \tag{19}$$

Hence, the channels $\mathcal{F}_\pm$ are completely determined by the transition probabilities $p_\pm(k|0)$ from $|0\rangle\langle 0|$ to the other energy eigenstates. Alternatively, one may say that the dynamics is identical along all of the diagonals (which does *not* imply that the collection of diagonals of the density matrix are identical).

Our primary goal in this section is to regain (1) from (2). However, we face an immediate obstacle in that (1) very explicitly refers to 'work', while we in our constructions deliberately have avoided to specify what the quantum version of work exactly is supposed to be. A way to test Crooks theorem in a macroscopic setting would be to measure the energy content in the energy source before and after the experiment, in order to see how much energy that has been spent. This macroscopic type of two-point measurements makes sense since we can expect a macroscopic source to be in states that would not be significantly disturbed by energy measurements (e.g. in the sense of pointer states and quantum Darwinism [85–87]). Recalling the discussion in section II E, the decoupling of the diagonal and off-diagonal modes of coherence can be regarded as yielding a weaker form of stability. In this case all coherences are blatantly annihilated, but the evolution of the diagonal distribution is unaffected by repeated energy measurements. Since our aim is to regain the classical Crooks relation, this appears as an acceptable form of stability, and it also seems reasonable to identify work with the energy loss in the reservoir, $w = E_n - E_m = s(n-m)$, for the initial measurement outcome E_n and the final outcome E_m. For these two-point measurements, the probability $P_\pm(w)$ is obtained by summing over all possible transitions that result in the work w for the given initial state σ, and thus

$$P_\pm(w) = \sum_{m,n:E_n-E_m=w} p_\pm(m|n)\langle n|\sigma|n\rangle. \tag{20}$$

To be able to treat the energy reservoir as an implicit object, as it is in the standard classical case, the distributions $P_\pm$ should not be functions of the initial state σ. Here, the energy translation invariance comes into play. By combining (20) with (18) we find that $P_\pm(w) = p_\pm(-w/s|0) = p_\pm(0|w/s)$, which thus removes the dependence on the reservoir. In the final step we combine the last equality with the diagonal Crooks relation (11) and obtain the classical Crooks relation (1).

III. CONDITIONAL FLUCTUATION THEOREMS

There are several reasons for why it is useful to generalize the type of quantum fluctuation theorems that we

have considered so far. First of all, the fluctuation relation (2) and its accompanying setup is in many ways an idealization. For example, the requirement of perfect control together with energy conservation is a strong assumption that easily leads to an energy reservoir that has to have an energy spectrum that is unbounded from both above and below (see Appendix C 4) and a bottomless spectrum is certainly not very reasonable from a physical point of view. A further consequence of the idealization, which may not be apparent unless one dives into the technicalities in the appendices, is that (2) is based on rather elaborate assumptions. Apart from resolving these issues, the conditional fluctuation theorems naturally incorporate non-equilibrium initial states. (There exist previous generalizations to non-canonical initial states. See e.g. [88] for a classical case, and [18, 25, 78, 89] for the quantum case.)

To see how the conditional fluctuation relations can come about, let us again consider the perfect control mechanism, i.e., that the evolution transforms the initial control state into the final with certainty. Imagine now that we abandon this assumption, and instead include a measurement that checks whether the control system succeeds to reach the final control state or not. The idea is that we only accept the particular run of the experiment if the control measurement is successful. Analogous to our previous fluctuation relations, the conditional fluctuation theorem relates the induced dynamics of the forward and reverse processes, but which now are conditioned on successful control measurements (see Fig. 5).

It is useful to keep in mind that one should be cautious when interpreting post-selected dynamics, as it easily can end up in seemingly spectacular results. The conditional dynamics can for example create states with off-diagonal elements from globally diagonal states (see Appendix H 7 b for an explicit example). However, this should not viewed as a mysterious creation of coherence, but simply as the result of a post selection with respect to a non-diagonal measurement operator.

In the introduction it was pointed out that an advantage with fluctuation theorem (2) is that it does not rely on any auxiliary measurements. It may thus appear a bit contradictory that we here re-introduce measurements as a fundamental component in the formalism. However, these only serve as control measurements upon which we condition the quantum evolution in the energy reservoir. This is in contrast to approaches where the purpose of the measurements is to generate classical outcomes (as for two-point energy measurements [8, 11–25]) and where the quantum fluctuation relations are based on the resulting probability distributions, much in the spirit of classical fluctuation relations.

As a final general remark one should note that it turns out to be unnecessary to restrict the control measurements to the control system; we can let them act on the joint system $\tilde{S} = S'C = SBC$ in any manner that we wish, resulting in a conditional evolution on E.

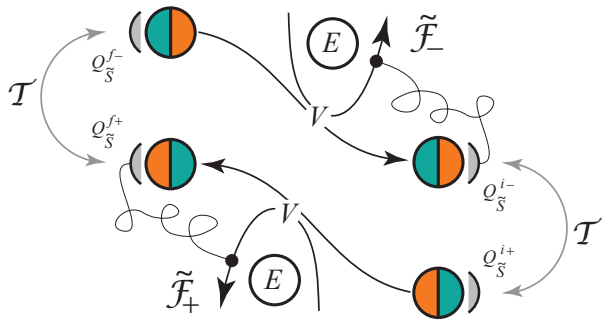

FIG. 5: **Conditional fluctuation theorems.** Similar to our previous fluctuation theorems we here relate the dynamics induced on the energy reservoir E by the forward and reverse process, but conditioned on successful control measurements on $\tilde{S} = SBC$. The forward process is characterized by a pair of positive semi-definite operators $Q^{i+}_{\tilde{S}}, Q^{f+}_{\tilde{S}}$ on $\tilde{S}$, such that the initial state is given via the Gibbs map $\mathcal{G}_{\beta H_{\tilde{S}}}(Q^{i+}_{\tilde{S}})$, and the control measurement is represented by $Q^{f+}_{\tilde{S}}$. More precisely, $Q^{f+}_{\tilde{S}}$ corresponds to the 'successful' outcome in the POVM $\{Q^{f+}_{\tilde{S}}, \hat{1}_{\tilde{S}} - Q^{f+}_{\tilde{S}}\}$, and results in the induced CPM $\tilde{\mathcal{F}}_+$ on the reservoir. The reverse process, resulting in the CPM $\tilde{\mathcal{F}}_-$, is correspondingly characterized by the pair $Q^{f-}_{\tilde{S}} = \mathcal{T}_{\tilde{S}}(Q^{f+}_{\tilde{S}})$, $Q^{i-}_{\tilde{S}} = \mathcal{T}_{\tilde{S}}(Q^{i+}_{\tilde{S}})$, where $\mathcal{G}_{\beta H_{\tilde{S}}}(Q^{f-}_{\tilde{S}})$ is the initial state, and $Q^{i-}_{\tilde{S}}$ gives the control measurement. The CPMs $\tilde{\mathcal{F}}_\pm$ are related via the conditional fluctuation theorem in (25). We are free to choose the measurement operators in any way we wish, and in the finite-dimensional case we can obtain arbitrary initial states. Hence, we are not restricted to initial equilibrium states. The price that we pay for this additional freedom is that non-equilibrium initial states are translated to non-trivial control measurements in the reverse process.

A. The Gibbs map and the partition map

Each control measurement has the two possible outcomes 'yes' or 'no', and can be described via measurement operators $0 \leq Q \leq \hat{1}$, where $\mathrm{Tr}(Q\rho)$ is the probability of a 'yes' when measured on a state ρ. Similarly $\mathrm{Tr}[(1-Q)\rho]$ is the probability of the outcome 'no'. In other words, $\{Q, \hat{1} - Q\}$ is a binary positive operator valued measure (POVM) [45, 90, 91]. These measurement operators do in our formalism serve a dual purpose; not only describing control measurements, but also parameterizing initial states. The latter is done via what we here refer to as the 'Gibbs map' $\mathcal{G}_{\beta H}$ and the 'partition map' $\mathcal{Z}_{\beta H}$ defined by

$$\mathcal{G}_{\beta H}(Q) = \frac{1}{\mathcal{Z}_{\beta H}(Q)} \mathcal{J}_{\beta H}(Q), \quad \mathcal{Z}_{\beta H}(Q) = \mathrm{Tr}\mathcal{J}_{\beta H}(Q). \tag{21}$$

For finite-dimensional Hilbert spaces all density operators can be reached via $\mathcal{G}_{\beta H}$ for suitable choices of Q. These mappings are generalizations of the Gibbs state and the partition function in the sense that $\mathcal{G}_{\beta H}(\hat{1}) =$

$G(H)$ and $\mathcal{Z}_{\beta H}(\hat{1}) = Z(H)$.

B. A conditional fluctuation relation

As before, we assume a global non-interacting Hamiltonian of the form

$$H = H_{\widetilde{S}} \otimes \hat{1}_E + \hat{1}_{\widetilde{S}} \otimes H_E. \tag{22}$$

We furthermore assume time-reversals $\mathcal{T}_{\widetilde{S}}$ and $\mathcal{T}_E$ such that $\mathcal{T}_{\widetilde{S}}(H_{\widetilde{S}}) = H_{\widetilde{S}}$, and $\mathcal{T}_E(H_E) = H_E$. We also assume that the global evolution preserves energy $[V, H] = 0$, and that it is time-reversal symmetric, $\mathcal{T}(V) = V$, with respect to $\mathcal{T} = \mathcal{T}_{\widetilde{S}} \otimes \mathcal{T}_E$.

The forward process is characterized by a pair of measurement operators on system $\widetilde{S}$, where $Q^{i+}_{\widetilde{S}}$ defines the initial state via the Gibbs mapping, and $Q^{f+}_{\widetilde{S}}$ the measurement at the end of the process. (These measurement operators do not have to be related to energy measurements.) The global initial state $\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{i+}_{\widetilde{S}}) \otimes \sigma$ evolves via an energy-preserving unitary operation V, and we condition the result on a successful measurement of $Q^{f+}_{\widetilde{S}}$. The resulting completely positive map (CPM) on E is given by

$$\tilde{\mathcal{F}}_+(\sigma) = \mathrm{Tr}_{\widetilde{S}}\big([Q^{f+}_{\widetilde{S}} \otimes \hat{1}_E] V [\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{i+}_{\widetilde{S}}) \otimes \sigma] V^\dagger\big). \tag{23}$$

That we generically get a CPM rather than a channel corresponds to the fact that the control measurement may fail. The quantity $\mathrm{Tr}\tilde{\mathcal{F}}_+(\sigma)$ is the probability that the control measurement succeeds.

To obtain the reversed process we define the operators $Q^{i-}_{\widetilde{S}} = \mathcal{T}_{\widetilde{S}}(Q^{i+}_{\widetilde{S}})$ and $Q^{f-}_{\widetilde{S}} = \mathcal{T}_{\widetilde{S}}(Q^{f+}_{\widetilde{S}})$, and use $Q^{f-}_{\widetilde{S}}$ to generate the initial state, and $Q^{i-}_{\widetilde{S}}$ to characterize the final measurement. In other words, we not only time-reverse the measurement operators, but also swap their roles. The resulting CPM on E for the reversed process thus becomes

$$\tilde{\mathcal{F}}_-(\sigma) = \mathrm{Tr}_{\widetilde{S}}\big([Q^{i-}_{\widetilde{S}} \otimes \hat{1}_E] V [\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{f-}_{\widetilde{S}}) \otimes \sigma] V^\dagger\big). \tag{24}$$

With the above assumptions one can show (see Appendix H) that the CPMs $\tilde{\mathcal{F}}_\pm$ are related by the following 'conditional' fluctuation theorem

$$\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}})\tilde{\mathcal{F}}_+ = \mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}})\mathcal{J}_{\beta H_E}\tilde{\mathcal{F}}^{\ominus}_-\mathcal{J}^{-1}_{\beta H_E}, \tag{25}$$

where we use the notation $\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}}) = \mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^{i\pm}_{\widetilde{S}})$ and $\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}}) = \mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^{f\pm}_{\widetilde{S}})$.

The conditional fluctuation relation (25) allows (up to infinite-dimensional technicalities) for arbitrary initial states on $\widetilde{S}$, and thus in particular arbitrary coherences. We only need to choose the appropriate initial measurement operator $Q^{i+}_{\widetilde{S}}$ that generates the desired initial state via the Gibbs map.

The quantum fluctuation relation in (2) can be regained from (25), and as a first step we choose the measurement operators to be $Q^{i\pm}_{\widetilde{S}} = \hat{1}_{S'} \otimes |c_{i\pm}\rangle\langle c_{i\pm}|$ and $Q^{f\pm}_{\widetilde{S}} = \hat{1}_{S'} \otimes |c_{f\pm}\rangle\langle c_{f\pm}|$. For the appropriate Hamiltonian $H_{\widetilde{S}}$, the Gibbs map gives the conditional equilibrium states $\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{i\pm}_{\widetilde{S}}) = G(H^i_{S'}) \otimes |c_{i\pm}\rangle\langle c_{i\pm}|$ and $\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{f\pm}_{\widetilde{S}}) = G(H^f_{S'}) \otimes |c_{i\pm}\rangle\langle c_{i\pm}|$. Due to the additional assumption of perfect control, the evolution V is such that we can replace the final measurement operators with identity operators, i.e., we do not have to perform any measurements at the end of the processes.

This sketchy re-derivation indicates that we can abolish the assumption of perfect control if we instead keep the measurements at the end of the protocol. Hence, these control measurements may fail, e.g., if the energy reservoir does not contain sufficient energy, which thus allows us to avoid an energy spectrum that is unbounded from below. (For an explicit example, see Appendix H 7 a.)

As opposed to (2) there is for (25) in general no decomposition into different off-diagonal Crooks relations. However, in the special case that the measurement operators $Q^{i+}_{\widetilde{S}}$ and $Q^{f+}_{\widetilde{S}}$ are diagonal with respect to an eigenbasis of $H_{\widetilde{S}}$ one can obtain (see Appendix H 6) diagonal conditional versions. In particular, one can impose the energy translation invariance and obtain (Appendix H 6) a conditional version of the classical Crooks relation.

IV. AN ALTERNATIVE FORMULATION

So far we have formulated our fluctuation relations solely in terms of channels and CPMs induced on the energy reservoir. Here we consider a reformulation that highlights two elementary properties on which our fluctuation relations in some sense are based.

Given some channel or CPM $\mathcal{F}$, a pair of measurement operators Q^i and Q^f, and a Hamiltonian H, we define the 'transition probability' from Q^i to Q^f by

$$P^{\mathcal{F}}_{\beta H}[Q^i \to Q^f] = \mathrm{Tr}\Big(Q^f \mathcal{F}\big(\mathcal{G}_{\beta H}(Q^i)\big)\Big). \tag{26}$$

Hence, $P^{\mathcal{F}}_{\beta H}[Q^i \to Q^f]$ is the probability that we would detect Q^i after we have evolved the initial state $\mathcal{G}_{\beta H}(Q^i)$ under the map $\mathcal{F}$. The Gibbs map thus again appears as a method to parametrize the set of initial states.

The quantum Crooks relation (2) can be rephrased in terms of these transition probabilities as

$$\begin{aligned} &Z(H^i_S)\mathcal{Z}_{\beta H_E}(Q^i_E)P^{\mathcal{F}_+}_{\beta H}[Q^{i+}_E \to Q^{f+}_E] \\ &= Z(H^f_S)\mathcal{Z}_{\beta H_E}(Q^f_E)P^{\mathcal{F}_-}_{\beta H}[Q^{i-}_E \to Q^{f-}_E], \end{aligned} \tag{27}$$

where Q^{i+}_E, Q^{f+}_E are measurement operators on the energy reservoir (not on $\widetilde{S}$), and where, as one may expect, $Q^{i-}_E = \mathcal{T}_E(Q^{i+}_E)$ and $Q^{f-}_E = \mathcal{T}_E(Q^{f+}_E)$.

In section II D we noted the possibility to interpret the quantum Crooks theorem (2) in terms of process tomography of the two induced channels $\mathcal{F}_\pm$. In that interpretation the mapping $\mathcal{F}_- \mapsto \mathcal{J}_{\beta H_E}\mathcal{F}_-^{\ominus}\mathcal{J}_{\beta H_E}^{-1}$ is something that we implement 'by hand' on the experimentally determined description of the channel $\mathcal{F}_-$. The reformulation (27) suggests a different interpretation where the fluctuation relation can be tested more directly via the estimated transition probabilities. The mappings $\mathcal{J}_{\beta H_E}$ and $\ominus$ are in some sense put in by hand also in this scenario, but they enter in the parametrization of the initial states and the measurement operators, rather than via a post processing of the measurement data.

It is no coincidence that the measurement operators $Q_E^{i\pm}, Q_E^{f\pm}$ undergo the same transformations as the operators $Q_{\widetilde{S}}^{i\pm}, Q_{\widetilde{S}}^{f\pm}$ that we know from the conditional fluctuation relations. These similarities stem from the fact that all of our fluctuation relations can be regarded as special cases of a 'global' fluctuation relation, defined for global measurement operators Q^{i+}, Q^{f+} on the whole of $SBCE$. It is straightforward to confirm (or to consult Appendix I) that energy conservation $[H, V] = 0$ combined with time reversal symmetry $\mathcal{T}(H) = H$, $\mathcal{T}(V) = V$, yields

$$\mathcal{Z}_{\beta H}(Q^i)\mathcal{P}_{\beta H}^V[Q^{i+} \to Q^{f+}] = \mathcal{Z}_{\beta H}(Q^f)\mathcal{P}_{\beta H}^V[Q^{f-} \to Q^{i-}], \tag{28}$$

where $Q^{f-} = \mathcal{T}(Q^{f+})$ and $Q^{i-} = \mathcal{T}(Q^{i+})$.

The global relation (28) can alternatively be phrased as the invariance of the quantity $\mathrm{Tr}\big(Q^f V \mathcal{J}_{\beta H}(Q^i) V^\dagger\big)$ under the transformation $(Q^i, Q^f) \mapsto (Q^{i'}, Q^{f'}) = \big(\mathcal{T}(Q^f), \mathcal{T}(Q^i)\big)$. All our fluctuation relations do in some sense express this basic invariance in different guises.

One can indeed re-derive (2) and (25) from (28). However, when doing so one quickly realizes that this hinges on another property (that we have used repeatedly without comments). Namely, if the global Hamiltonian H is non-interacting over two subsystems, i.e. $H = H_1 \otimes \hat{1}_2 + \hat{1}_1 \otimes H_2$, then the mapping $\mathcal{J}$ satisfies the factorization property

$$\mathcal{J}_{\beta H}(Q_1 \otimes Q_2) = \mathcal{J}_{\beta H_1}(Q_1) \otimes \mathcal{J}_{\beta H_2}(Q_2). \tag{29}$$

Although elementary, it is in some sense this property (in conjunction with product time-reversals $\mathcal{T} = \mathcal{T}_1 \otimes \mathcal{T}_2$) that makes it possible to formulate fluctuation relations solely in terms of the dynamics of the energy reservoir, and also to eliminate unaccessible degrees of freedom (see Appendix I for further details). The central role of this property may become more apparent when we abandon it in section VI .

As a side-remark one can note that if one attempts to formulate quantum fluctuation relations for non-Gibbsian generalizations of equilibrium distributions, then the lack of a factorization property for non-exponential functions makes the generalization problematic (see Appendix I 4).

V. CORRELATED INITIAL STATES

The conditional relations extend the notion of fluctuation theorems to non-equilibrium initials states on $\widetilde{S}$. An obvious question is if also entanglement and general pre-correlations can be included. (One could even argue that in order to deserve the label 'fully quantum' our fluctuation relations must include quantum correlations.) The global fluctuation relation (28) already provides an affirmative answer to this question, in the sense that we can generate arbitrary global initial states via suitable choices of measurement operators (including quantum correlations to external reference systems). However, (28) is not the only option for how to describe such scenarios, and here we highlight three special cases that focus on descriptions via channels or CPMs.

A. Correlations between E and an external reference

Imagine that the energy reservoir not only carries energy and coherences, but also carries entanglement and correlations with an external reference R (not included in $SBCE$). One option for obtaining a fluctuation relation that incorporates this scenario would be to regard ER as a new extended energy reservoir, with a Hamiltonian H_{ER} and a new time-reversal $\mathcal{T}_{ER}$, with $\mathcal{T}_{ER}(H_{ER}) = H_{ER}$. Both the quantum fluctuation relation (2) and the conditional fluctuation relation (25) are applicable to this scenario. In the latter case, the induced CPMs $\mathcal{F}_\pm$ on the new energy reservoir ER satisfies the corresponding quantum Crooks relation $\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q_{\widetilde{S}}^i)\mathcal{F}_+ = \mathcal{Z}_{\beta H_{\widetilde{S}}}(Q_{\widetilde{S}}^f)\mathcal{J}_{\beta H_{ER}}\mathcal{F}_-^{\ominus}\mathcal{J}_{\beta H_{ER}}^{-1}$.

B. Correlations between $\widetilde{S}$ and an external reference

An obvious alternative to the energy reservoir carrying correlations with an external reference is that $\widetilde{S}$ carries these correlations. We can analogously incorporate R into $\widetilde{S}$, with the joint Hamiltonian $H_{\widetilde{S}R}$ and an extended time-reversal $\mathcal{T}_{\widetilde{S}R}$ such that $\mathcal{T}_{\widetilde{S}R}(H_{\widetilde{S}R}) = H_{\widetilde{S}R}$. The requirements for the conditional fluctuation theorem (25) are satisfied for this extended system, thus resulting in the relation $\mathcal{Z}_{\beta H_{\widetilde{S}R}}(Q_{\widetilde{S}R}^i)\tilde{\mathcal{F}}_+ = \mathcal{Z}_{\beta H_{\widetilde{S}R}}(Q_{\widetilde{S}R}^f)\mathcal{J}_{\beta H_E}\tilde{\mathcal{F}}_-^{\ominus}\mathcal{J}_{\beta H_E}^{-1}$.

C. Pre-correlated $\widetilde{S}$ and E

The two previous examples can both be regarded as variations of the conditional fluctuation relation, and do in this sense not add anything essentially new to the general picture. A maybe more interesting case is if we allow

for pre-correlations between $\widetilde{S}$ and E. (It is not difficult to imagine cases where the energy reservoir E interacts repeatedly with $\widetilde{S}$, and thus may build up correlations.) Apart from the global fluctuation relation (28) this case does not fit very well with our previous scenarios. However, it is straightforward to adapt our general formalism to find a fluctuation relation for the evolution on the joint system SE.

For the sake of illustration we consider one particular case related to the setup in section II D, but where we allow for initial correlations between S and E. In essence (for details on the setup, see Appendix J) we have an initial Hamiltonian H^i_{SE} and a final Hamiltonian H^f_{SE} and assume perfect control that transfers one into the other, resulting in the induced channels

$$\begin{aligned}\overline{\mathcal{F}}_+(\chi) =& \mathrm{Tr}_{CB}(V[|c_{i+}\rangle\langle c_{i+}| \otimes G(H_B) \otimes \chi]V^\dagger),\\ \overline{\mathcal{F}}_-(\chi) =& \mathrm{Tr}_{CB}(V[|c_{f-}\rangle\langle c_{f-}| \otimes G(H_B) \otimes \chi]V^\dagger)\end{aligned}$$

on the combined system SE. For these channels one can derive the fluctuation relation

$$\overline{\mathcal{F}}_+ = \mathcal{J}_{\beta H^f_{SE}} \overline{\mathcal{F}}^{\ominus}_- \mathcal{J}^{-1}_{\beta H^i_{SE}}. \tag{30}$$

The partition maps on the left and right hand side cancel due to the identical initial and final Hamiltonian for a heat bath that starts in the Gibbs state. Note also that the two applications of the $\mathcal{J}$-map may potentially involve an initial Hamiltonian H^i_{SE} that is different from the final Hamiltonian H^f_{SE}.

VI. APPROXIMATE FLUCTUATION RELATIONS

The global Hamiltonian H has so far in this investigation only characterized the notion of energy conservation, while the dynamics has been modeled by a unitary operator V, with the only restriction that H and V should commute. In view of standard textbook quantum mechanics it would be very reasonable to demand a much more tight connection, where the global Hamiltonian H induces the evolution according to Schrödinger's equation, thus yielding $V = e^{-itH/\hbar}$. An additional benefit with the latter arrangement would be that it does not require any further interventions beyond the preparation of the initial state, the measurement of the final state, and the time-keeping for when to do the measurement (which may require non-trivial resources). In other words, once we have started the global system it evolves autonomously. (For discussions on autonomous and clock controlled thermal machines in the quantum regime, see [92, 93].) This should be compared with the models that we have employed so far, where one in principle should analyze the mechanism that implements the evolution V.

The problem is that if we would impose the condition $V = e^{-itH/\hbar}$ for non-interacting Hamiltonians as in (22), the resulting dynamics on the energy reservoir would become trivial. If we on the other hand would abandon the non-interacting Hamiltonians, then we also have to abandon the factorization property (29), which, as pointed out in section IV, plays an important role in our derivations. The main observation in this section is that we can obtain approximate fluctuation relations as long as the factorization holds approximately, which also is enough to obtain a non-trivial evolution. In the following section we shall illustrate the general ideas with two special cases. (For a more general version that includes both of them, see Appendix K.)

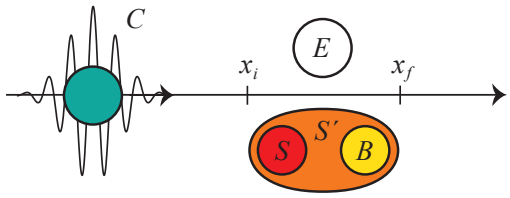

FIG. 6: **A control particle.** One way to 'quantize' the control parameter x of the classical Crooks relation would be to regard it as the position of a quantum particle. The propagation of the particle would approximately implement the time dependent Hamiltonian for suitable wave packets. A particularly clean example is obtained if E and S' only interact when the control particle is in an 'interaction region' $[x_i, x_f]$ (cf. the construction in [92]). Since the global Hamiltonian contains interactions between E and S' it does not fit with our previous classes of models, and in particular, it does not satisfy the factorization property (29). However, for suitable choices of measurement operators Q^{i+}_C and Q^{f+}_C that are localized outside the interaction region, the systems becomes approximately non-interacting, and the the factorization property (29) is approximately satisfied. This makes it possible to derive an approximate conditional fluctuation relation on the energy reservoir.

A. Approximate conditional fluctuation relations

There is an additional reason for why it is useful to go beyond our previous settings. Namely that there exist rather evident ways to 'quantize' the classical control mechanism in Crooks relation that do not fit particularly well within the machinery that we have employed so far. These quantized control mechanisms moreover provide a good starting point for introducing approximate fluctuation relations.

Imagine a particle whose position x determines the parameter in the family of Hamiltonians $H_{S'E}(x)$. More precisely, assume a joint Hamiltonian of the form

$$H = \frac{1}{2M_C}\hat{P}^2_C \otimes \hat{1}_{S'E} + H_{S'E}(\hat{X}_C), \tag{31}$$

where M_C is the mass, and $\hat{X}_C, \hat{P}_C$ are the canonical position and momentum operators of the control particle. If the control particle is reasonably well localized in both space and momentum, one can imagine that its

propagation approximately implements the evolving control parameter x in the family of Hamiltonians $H_{S'E}(x)$. For the sake of simplicity, let us assume that the control particle only affects the other systems in an 'interaction region' (see Fig. 6.) corresponding to an interval $[x_i, x_f]$, while outside of this region it is the case that

$$H_{S'E}(x) = \begin{cases} H^i_{S'} \otimes \hat{1}_E + \hat{1}_{S'} \otimes H_E, & x \leq x_i, \\ H^f_{S'} \otimes \hat{1}_E + \hat{1}_{S'} \otimes H_E, & x \geq x_f. \end{cases} \tag{32}$$

Although the systems thus are non-interacting outside the interaction region $[x_i, x_f]$ one should keep in mind that the kinetic energy term $\hat{K} = \hat{P}^2_C/(2M_C)$ in (31) does not commute with X_C, and thus prohibits a factorization as in (29). In some sense it is thus the kinetic operator $\hat{K}$ that causes the failure of the factorization property. On the other hand, it is also the kinetic operator that yields a non-trivial evolution of the control mechanism. It is this tension that we strive to handle via the approximate fluctuation relations. The basic idea is that if the measurement operators Q^{i+}_C and Q^{f+}_C are well localized outside the interaction region, then the systems are approximately non-interacting, and consequently the factorization (29) should hold approximately. It is worth keeping in mind that even if the measurement operators are well localized in non-interacting regions, this does not exclude the possibility that the system evolves into states where the interactions are strong, even at the very moment of the control measurement. (For an explicit example, see Fig. 10 in Appendix K 3 c.)

Although the special case of the control particle provides intuition, the more general setting of the conditional fluctuation theorems yields a more concise description. For the chosen measurement operators $Q^{i\pm}_{\tilde{S}}$ and $Q^{f\pm}_{\tilde{S}}$, let us assume that there exist local approximate Hamiltonians $H^i_{\tilde{S}}$, $H^f_{\tilde{S}}$, H^i_E, H^f_E, such that the factorization holds approximately

$$\begin{aligned} \mathcal{J}_{\beta H}(Q^{i+}_{\tilde{S}} \otimes Q) &\approx \mathcal{J}_{\beta H^i_{\tilde{S}}}(Q^{i+}_{\tilde{S}}) \otimes \mathcal{J}_{\beta H^i_E}(Q), \\ \mathcal{J}_{\beta H}(Q^{f+}_{\tilde{S}} \otimes Q) &\approx \mathcal{J}_{\beta H^f_{\tilde{S}}}(Q^{f+}_{\tilde{S}}) \otimes \mathcal{J}_{\beta H^f_E}(Q). \end{aligned} \tag{33}$$

For time-reversal symmetric systems one can show that this leads to the approximate fluctuation relation

$$\mathcal{Z}_{\beta H^i_{\tilde{S}}}(Q^{i+}_{\tilde{S}})\tilde{\mathcal{F}}_+ \mathcal{J}_{\beta H^i_E} \approx \mathcal{Z}_{\beta H^f_{\tilde{S}}}(Q^{f-}_{\tilde{S}})\mathcal{J}_{\beta H^f_E}\tilde{\mathcal{F}}^{\ominus}_-. \tag{34}$$

This can be turned into a quantitative statement where the size of the error in (34) is bounded by the errors in (33), see Appendix K 2.

B. Joint control system and energy reservoir

Compared to the control particle in Fig. 6 there exists an even simpler setup where the control particle simultaneously serves as the energy reservoir (see Fig. 7). This

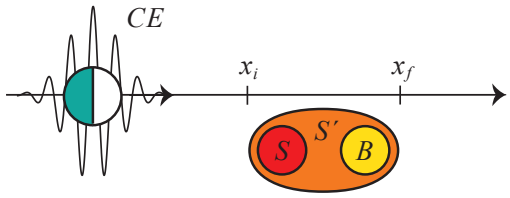

FIG. 7: **A control particle that also serves as energy reservoir.** An alternative to the setup in Fig. 6 is a particle that simultaneously serves as both control system and energy reservoir. Hence, it is the kinetic energy of the control particle that drives the non-equilibrium process (if the other systems would start in equilibrium). This is another example of a system that would generally not satisfy the factorization property (29), but for measurement operators Q^{i+}_{CE} and Q^{f+}_{CE} that are localized outside the interaction region, the systems becomes approximately non-interacting. For this setup one can obtain an approximate fluctuation relation in terms of the transition probabilities introduced in section IV.

corresponds to the global Hamiltonian

$$H = \frac{1}{2M_{CE}}\hat{P}^2_{CE} \otimes \hat{1}_{S'} + H_{S'}(\hat{X}_{CE}), \tag{35}$$

again with an interaction region

$$H_{S'}(x) = \begin{cases} H^i_{S'}, & x \leq x_i, \\ H^f_{S'}, & x \geq x_f, \end{cases} \tag{36}$$

with a non-trivial dependence on x inside $[x_i, x_f]$. As opposed to the previous setting, we here perform the control measurement on the energy reservoir itself. This situation is conveniently described in terms of the transition probabilities discussed in section IV, resulting in the approximate fluctuation relation

$$\begin{aligned} &\mathcal{Z}_{\beta H^i_{S'}}(Q^i_{S'})\mathcal{Z}_{\beta H^i_{CE}}(Q^i_{CE}) \\ &\quad \times P^V_{\beta H^i}[Q^{i+}_{S'} \otimes Q^{i+}_{CE} \to Q^{f+}_{S'} \otimes Q^{f+}_{CE}] \\ &\approx \mathcal{Z}_{\beta H^f_{S'}}(Q^f_{S'})\mathcal{Z}_{\beta H^f_{CE}}(Q^f_{CE}) \\ &\quad \times P^V_{\beta H^f}[Q^{f-}_{S'} \otimes Q^{f-}_{CE} \to Q^{i-}_{S'} \otimes Q^{i-}_{CE}]. \end{aligned} \tag{37}$$

where for the particular choice of Hamiltonian (35) we have $H^i_{CE} = H^f_{CE} = \hat{P}^2_{CE}/(2M_{CE})$. This approximate relation can also be made quantitative, see Appendix K 3. For a numerical evaluation of the errors in a concrete model, see Appendix K 3 c.

VII. CONCLUSIONS AND OUTLOOK

We have generalized Crooks fluctuation theorem to a genuine quantum regime that incorporates the full quantum dynamics. This leads to a decomposition into diagonal and off-diagonal Crook's relations, one for each mode of coherence. We have also derived Jarzynski equalities, and re-derived a standard bound on the average work

cost under an additional assumption of unitality of a certain induced channel. We have furthermore shown that the classical Crooks relation can be regained under the additional assumption of energy translation invariant dynamics on the energy reservoir. The general approach moreover leads to the concept of conditional fluctuation relations, where a pair of measurement operators characterizes the initial state and the final measurement, and where the transformation from the forward to the reverse process corresponds to a transformation of the pair of measurement operators. This generalization allows for non-equilibrium initial states, and can also be extended to include correlations. Finally we have demonstrated that by allowing for errors in the fluctuation relation, we can incorporate the 'natural' setting where the global dynamics is determined by a single time-independent Hamiltonian.

One can speculate if it would be possible to use the latter global Hamiltonian as a starting point to obtain approximate fluctuation relations for master equations by reduction of the heat bath along the lines of standard derivations (see e.g. [94]). This may yield models that would be more amenable to numerical modeling. In this context one should note previous approaches to fluctuation relations for master equations [95, 96], unravellings [97], and Brownian motion models [98].

Classical fluctuation relations have been subject to several experimental tests [99–104]. Various setups have been suggested for the quantum case [105–111], with recent experimental implementations in NMR [112], as well as in a trapped ion systems [113]. See also [114] for an experiment on a relation for non-thermal noise. The conditional fluctuation theorems, and in particular the approximate version, allows for a considerable flexibility, which suggest that experimental tests may be feasible. Since the quantum fluctuation relation (2) and the conditional version (25) are phrased in therms of channels and CPMs, and one may be tempted to conclude that every test of these relations would necessarily require a full process tomography, which generally is very demanding. However, the global fluctuation relation (28) suggests 'milder' tests based on small sets of suitable chosen measurement operators (corresponding to a partial process tomography).

One could also consider the possibility to experimentally verify cases of Jarzynski equalities with and without the condition of unitality of the channel $\mathcal{R}_+$, such as in (13) and (14). It would also be desirable to get a better theoretical understanding of the role of the unitality of the channel $\mathcal{R}_+$, which one may suspect is related to the energy reservoir regarded as a resource. In this context one may also ask for the general conditions for the non-violation of standard bounds, and how this relates to the energy translation invariance (cf. discussions in [44]).

In this investigation we have tacitly assumed that the heat bath can be taken as initially being in the Gibbs state. It would be desirable to let go of this assumption, e.g. via typicality [115, 116] (see further discussions in Appendix L).

On a more technical note one may observe that whenever we actively have referred to the properties of the spectrum of the Hamiltonian of the reservoir, we have always assumed that it is a pure point spectrum. An analysis that explicitly investigates the effects of reservoir spectra that contains a continuum could potentially be useful.

One can imagine several generalizations of the results in this investigation. It does for example seem plausible with a grand canonical version, thus not only including energy flows, but also the flow of particles. (For a previous grand canonical fluctuation relation see [117].)

Another potential generalization concerns a classical version of the conditional fluctuation relations. We have already obtained a particular class of classical conditional fluctuation theorems (Appendix H 6 a). However, these are classical in the sense of being diagonal with respect to a fixed energy eigenbasis, which should not be confused with a classical phase space setting. It seems reasonable that the structure of pairs of measurement operators, translated to functions over phase space, with classical counterparts for the Gibbs and partition maps, could combine with phase space flows to yield classical conditional fluctuation theorems. It may also be possible to bridge such a classical phase space approach to the quantum setting via Wigner functions and other phase space representations of quantum states.

The intermediate fluctuation relation (6) can be rephrased in terms of Petz recovery channel [47–50], which reminds of the recent finding in [40] that relates Petz recovery channel with work extraction. One may wonder if these results hint at a deeper relation. See further comments in Appendix A 8.

Several recent contributions to quantum thermodynamics has focused on resource theories, single-shot statistical mechanics, and coherence [26–43, 120–132]. (For general overviews on later developments in quantum thermodynamics, see [137, 138].) The fluctuation relations in this investigation are at their core statements about dynamics, rather than about resources. Nevertheless, one could consider to formulate quantitative characterizations of the evolution of resources in the spirit of Crooks theorem, and our fluctuation relations may serve as a starting point for such an analysis. In this context one should note recent efforts to link single-shot quantities and fluctuation theorems [133–136]. The fact that the present investigation is based on energy conserving dynamics and thus brings the notion of fluctuation theorems under the same umbrella as previous investigations on quantum thermodynamics and coherence [26–43] may further facilitate the merging of these subjects.

Acknowledgements

This work was supported by the Excellence Initiative of the German Federal and State Governments (Grants ZUK 43 & 81), and the DFG.

Appendix A: An intermediate quantum fluctuation theorem

1. Setting the stage

The standard classical Crooks theorem compares the probability distributions of the random work costs of a forward and reverse process where the system is driven by external fields. Often this external field is taken as a parameter x in a Hamilton function. The system is usually imagined to additionally interact with a heat bath of a given temperature. The time-schedule of the parameter x is implemented as a function x_t of time t, which runs from $t = 0$ to $t = T$. At $t = 0$ we assume that the initial system is in equilibrium with the heat bath. For the reverse process, the external parameter evolves as $x'_t := x_{T-t}$ for $t = 0$ to $t = T$. In other words, the time-schedule of the parameter is run in reverse. Again we assume that the system initially is in equilibrium with the heat bath, but now for the parameter value $x'_0 = x_T$. It is useful to keep in mind that these initial equilibrium distributions are conditioned on the value of the control parameter. The aim of the following sections is to make a quantum version of this classical setup.

The model consists of four components, the 'system' S, the heat bath B, the control C, and the energy reservoir E. Assumptions 1 below does not mention the system S or the heath bath B. The reason for this is that the main part of the derivations does not require any distinction between these subsystems, so they can be regarded as one single system $S' := SB$.

For a Hamiltonian H and $\beta = 1/(kT)$, for Boltzmann's constant k and the absolute temperature T, we denote the partition function by $Z_\beta(H) = \mathrm{Tr}e^{-\beta H}$, and (assuming that $Z_\beta(H)$ is finite) we denote the Gibbs state by $G_\beta(H) = e^{-\beta H}/Z_\beta(H)$. Since we here only consider heat baths with one single temperature, we will often suppress the subscript and write $G(H)$ and $Z(H)$.

Assumptions 1. *Let $\mathcal{H}_{S'}$, $\mathcal{H}_C$, and $\mathcal{H}_E$ be complex Hilbert spaces. Let $|c_i\rangle, |c_f\rangle \in \mathcal{H}_C$ be normalized and orthogonal to each other, and define the projector $P_C^\perp := \hat{1}_C - |c_i\rangle\langle c_i| - |c_f\rangle\langle c_f|$.*

- *Let $H^i_{S'}$ and $H^f_{S'}$ be Hermitian operators on $\mathcal{H}_{S'}$, such that $Z_\beta(H^i_{S'})$ and $Z_\beta(H^f_{S'})$ are finite. (This guarantees that $G_\beta(H^i_{S'})$ and $G_\beta(H^f_{S'})$ exist.) Let H_E be a Hermitian operator on $\mathcal{H}_E$. Let $H^\perp$ be a Hermitian operator on $\mathcal{H}_{S'} \otimes \mathcal{H}_C$ such that $[\hat{1}_{S'} \otimes P_C^\perp]H^\perp[\hat{1}_{S'} \otimes P_C^\perp] = H^\perp$, and define*

$$H_{S'C} := H^i_{S'} \otimes |c_i\rangle\langle c_i| + H^f_{S'} \otimes |c_f\rangle\langle c_f| + H^\perp$$

and $H := H_{S'C} \otimes \hat{1}_E + \hat{1}_{S'C} \otimes H_E$.

- *V is a unitary operator on $\mathcal{H}_{S'} \otimes \mathcal{H}_C \otimes \mathcal{H}_E$ such that $[V, H] = 0$, and*

$$V[\hat{1}_{S'} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E] = [\hat{1}_{S'} \otimes |c_f\rangle\langle c_f| \otimes \hat{1}_E]V. \quad \text{(A1)}$$

In the following we briefly discuss the rationale behind these assumptions.

The Hamiltonian $H_{S'C}$ describes how the state of the control system C changes the Hamiltonian of S' (see Fig. 8). Since $|c_i\rangle$ is orthogonal to $|c_f\rangle$, and these in turn are orthogonal to the support of $H^\perp$, it follows that if C is in state $|c_i\rangle$, then the Hamiltonian of S' is $H^i_{S'}$. Similarly, if C is in state $|c_f\rangle$, then S' has Hamiltonian $H^f_{S'}$. The Hamiltonian $H^\perp$ allows for the possibility of having intermediate Hamiltonians between the initial and final one (see Fig. 8). The global Hamiltonian is the sum of $H_{S'C}$ and the Hamiltonian H_E of the energy reservoir, which thus by construction are non-interacting.

The global evolution is given by unitary operations that conserve energy, which here is modeled via unitary operators V on $S'CE$ such that $[H, V] = 0$. (For an alternative notion of energy conservation, see [44].) In addition to being energy conserving, we also require V to satisfy (A1). In other words, V should rotate the subspace $\mathcal{H}_{S'E} \otimes \mathrm{Sp}\{|c_i\rangle\}$ to the subspace $\mathcal{H}_{S'E} \otimes \mathrm{Sp}\{|c_f\rangle\}$. This models the idealization of a perfect control mechanism, meaning that the evolution with certainty will bring the initial control state $|c_i\rangle$ to the final control state $|c_f\rangle$, and thus with certainty will transform the initial Hamiltonian $H^i_{S'}$ to the final Hamiltonian $H^f_{S'}$. In section G 3 a we demonstrate that there exist setups that satisfy all conditions in Assumptions 1.

As mentioned above, we do not need to make a distinction between the system S and the heat bath B in most of these derivations. However, to obtain fluctuation relations where the partition functions only refer to system S, we can additionally assume that the initial and final Hamiltonians of the system and the heat bath are non-interacting. More precisely, we would assume that there exist Hermitian operators H^i_S and H^f_S on $\mathcal{H}_S$ and a Hermitian operator H_B on $\mathcal{H}_B$ such that

$$\begin{aligned} H^i_{S'} &= H^i_S \otimes \hat{1}_B + \hat{1}_S \otimes H_B, \\ H^f_{S'} &= H^f_S \otimes \hat{1}_B + \hat{1}_S \otimes H_B, \end{aligned} \quad \text{(A2)}$$

and such that $Z_\beta(H^i_S)$, $Z_\beta(H^f_S)$, and $Z_\beta(H_B)$ are finite.

2. Global Hamiltonian H and global evolution V

In typical textbook quantum mechanics the Hamiltonian defines the notion of energy and energy conservation, as well as being the generator of time-evolution. Here we do in some sense separate these two roles, since we let the time evolution be given by V with the only restriction that it commutes with H, without demanding that $V = e^{-itH/\hbar}$. This separation is very convenient since it gives tractable models (compared to introducing an interaction term in the Hamiltonian and try to analyze the resulting evolution via Schrödinger's equation) and has successfully been employed in several previous studies [26–43].

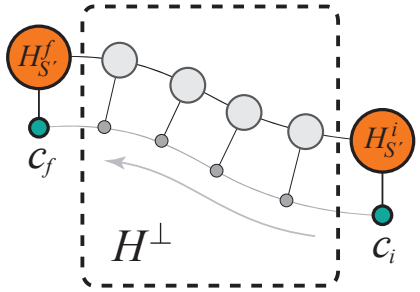

FIG. 8: **Structure of the Hamiltonian.** The Hamiltonian of the extended system $S' = SB$ and the control has the form $H_{S'C} = H^i_{S'} \otimes |c_i\rangle\langle c_i| + H^f_{S'} \otimes |c_f\rangle\langle c_f| + H^\perp$. Here $H^i_{S'}$ and $H^f_{S'}$ are the initial and final Hamiltonians, respectively, and $|c_i\rangle, |c_f\rangle$ the corresponding orthonormal control states. If the control is in state $|c_i\rangle$ the Hamiltonian of S' is $H^i_{S'}$, while it is $H^f_{S'}$ if the control is in state $|c_i\rangle$. The Hamiltonian $H^\perp$, which has orthogonal support to $H^i_{S'} \otimes |c_i\rangle\langle c_i|$ and $H^f_{S'} \otimes |c_f\rangle\langle c_f|$, corresponds to possible intermediate stages. In the proofs of our fluctuation theorems $H^\perp$ plays no particular role. However, it can be used to simulate a path of Hamiltonians $H_{S'}(x)$, e.g., via a discretization (see section G 3 b).

It is maybe worth emphasizing that when we in this investigation refer to two systems as being 'non-interacting', this only means that the energy observable is of the form $H = H_1 \otimes \hat{1}_2 + \hat{1}_1 \otimes H_2$. It does *not* imply that the evolution is of a trivial product form $V_1 \otimes V_2$, since (depending on the combination of the spectra of H_1 and H_2) there may be non-trivial unitary operators V that commute with H.

One way to understand the separation of roles between H and V is to imagine that the global evolution is generated by a Hamiltonian H_{evol}, i.e., that $V = e^{-itH_{\text{evol}}/\hbar}$, where we let $H_{\text{evol}} = H + H'$, $[H, H'] = 0$. A possible justification would be if H' is 'small', thus leaving H as the dominant contribution to the energy. It should be emphasized that the derivations of our fluctuation relations do not require us to know how V comes about, or what happens at intermediate times when the system evolves from the initial to the final state. We only need to know that V commutes with H.

Although the above reasoning may serve as a possible justification, one may nevertheless wonder how to incorporate more 'standard' assumption that $V = e^{-itH/\hbar}$. This topic is discussed in section K.

3. The initial states

In the typical derivation of Crooks theorem one assumes that the system initially is at equilibrium with respect to the initial value of the control parameter. In other words, for $x = x_i$, the system should start in state $G_\beta\big(H_{S'}(x_i)\big)$. When one considers a more explicit model that includes the degrees of freedom of the control system it becomes clear that the initial state of the system and control combined cannot be in a global equilibrium. For example, in our case the global equilibrium state would be

$$\begin{aligned} G(H_{S'C}) = &\frac{Z(H^i_{S'})}{Z(H_{S'C})} G(H^i_{S'}) \otimes |c_i\rangle\langle c_i| \\ &+ \frac{Z(H^f_{S'})}{Z(H_{S'C})} G(H^f_{S'}) \otimes |c_f\rangle\langle c_f| \\ &+ \frac{Z(H^\perp)}{Z(H_{S'C})} G(H^\perp), \end{aligned} \tag{A3}$$

where $Z(H_{S'C}) = Z(H^i_{S'}) + Z(H^f_{S'}) + Z(H^\perp)$. Hence, the global equilibrium is a weighted average over all the control states and the corresponding conditional equilibrium states in S'. We rather have to think of system S' as being in a 'conditional' equilibrium $G_\beta(H^i_{S'}) \otimes |c_i\rangle\langle c_i|$. The conditional equilibrium corresponds to a projection (and subsequent normalization) of the global equilibrium onto an eigenspace of $H_{S'C}$.

The initial state of the forward process is the conditional equilibrium state of $S'C$ and an arbitrary state σ of the reservoir, i.e., $G_\beta(H^i_{S'}) \otimes |c_i\rangle\langle c_i| \otimes \sigma$. In an analogous fashion, the reverse process should start in a conditional equilibrium with respect to the final Hamiltonian, thus corresponding to the global state $G_\beta(H^f_{S'}) \otimes |c_f\rangle\langle c_f| \otimes \sigma$.

The fact that we here assume that E and $S'C$ initially are uncorrelated makes it possible to formulate our quantum fluctuation theorems in terms of quantum channels on the energy reservoir alone.

4. Induced channels on the energy reservoir

For the standard formulations of Crooks theorem, the change of the external control parameters would typically push the system out of equilibrium at the expense of work. In our quantum treatment we wish to describe all aspects of how the state of the energy reservoir changes, which conveniently can be captured by the channels (trace preserving completely positive maps [45]) induced on the reservoir.

More precisely, we wish to describe how the state of the energy reservoir evolves under the action of a global energy conserving unitary operation V that additionally exhibits perfect control (A1). We furthermore assume that S' starts in the conditional equilibrium with respect to the initial control state $|c_i\rangle$ as described in the previous section. The state of the reservoir after the evolution can thus be written

$$\mathcal{F}(\sigma) := \mathrm{Tr}_{S'C}(V[G_\beta(H^i_{S'}) \otimes |c_i\rangle\langle c_i| \otimes \sigma]V^\dagger). \tag{A4}$$

Hence, $\mathcal{F}$ describes the change of state induced on the energy reservoir E due to the global dynamics V for this particular class of initial states.

For this intermediate version we reverse the entire evolution on the global system. More precisely, we replace $V \cdot V^\dagger$ with $V^\dagger \cdot V$. For a V generated by Hamiltonian evolution $V = e^{-itH_{\mathrm{evol}}/\hbar}$, this corresponds to a replacement of t with $-t$. The reverse process starts in the local equilibrium with respect to the final Hamiltonian $H^f_{S'}$, which results in the channel

$$\mathcal{R}(\sigma) := \mathrm{Tr}_{S'C}(V^\dagger[G_\beta(H^f_{S'}) \otimes |c_f\rangle\langle c_f| \otimes \sigma]V). \quad \text{(A5)}$$

Although this indeed guarantees that the evolution is reversed in a very concrete sense, one can argue that it does not quite correspond to the spirit of Crooks relation, which only requires a reversal of the control parameters. In Sec. C we will remove this idealization. The purpose of the following sections is to establish a relation (Proposition 1) between the 'forward' channel $\mathcal{F}$ and the 'reverse' channel $\mathcal{R}$.

5. Conjugate CPMs

The conjugate map ϕ^* of a completely positive map (CPM) ϕ can be defined via $\mathrm{Tr}\big(Y\phi(\sigma)\big) = \mathrm{Tr}\big(\phi^*(Y)\sigma\big)$, where Y are arbitrary (bounded) Hermitian operators, and σ arbitrary density operators. A convenient alternative characterization is via Kraus representations [45] $\phi(\sigma) = \sum_k V_k \sigma V_k^\dagger$, where the conjugate map is given by $\phi^*(Y) = \sum_k V_k^\dagger Y V_k$.

For the derivations it will be convenient to keep in mind the following observation. Suppose that a CPM ϕ is defined via a unitary $V : \mathcal{H}_a \otimes \mathcal{H} \to \mathcal{H}_a \otimes \mathcal{H}$ as

$$\phi(\sigma) := \mathrm{Tr}_a([Q_a \otimes \hat{1}]V[\eta_a \otimes \sigma]V^\dagger), \quad \text{(A6)}$$

where Q_a (bounded) and η_a (trace class) are positive operators on an ancillary Hilbert space $\mathcal{H}_a$. It follows that the conjugate CPM ϕ^* can be written

$$\phi^*(Y) = \mathrm{Tr}_a([\eta_a \otimes \hat{1}]V^\dagger[Q_a \otimes Y]V). \quad \text{(A7)}$$

6. The mapping $\mathcal{J}$

For an operator A we define the mapping

$$\mathcal{J}_A(Q) := e^{-A/2} Q e^{-A^\dagger/2}. \quad \text{(A8)}$$

The reason for why we here choose the exponent to be $-A/2$, rather than say A, is only to make it more directly related to Gibbs states in the special case that $A := \beta H$ and $Q := \hat{1}$, and thus $\mathcal{J}_{\beta H}(\hat{1}) = Z_\beta(H)G_\beta(H)$. The mapping $\mathcal{J}_A$ is a CPM, but is in general not trace preserving.

The mapping $\mathcal{J}_{\beta H}$ does often occur together with its inverse $\mathcal{J}^{-1}_{\beta H}$, in such a way that $\mathcal{J}_{\beta H} \circ \phi \circ \mathcal{J}^{-1}_{\beta H}$ for some CPM ϕ (see e.g. Proposition 1). This combination can in some sense be viewed as a quantum version of the term $e^{\beta w}$ in the classical Crooks relation in (1). To see this, let us consider the special case that H_E has a pure point spectrum, i.e., there exists an orthonormal basis of eigenvector $|n\rangle$ with corresponding eigenvalues E_n. For mappings from diagonal elements to diagonal elements we would get $\langle m|\mathcal{J}_{\beta H}\big(\phi\big(\mathcal{J}^{-1}_{\beta H}(|n\rangle\langle n|)\big)\big)|m\rangle = e^{\beta(E_n - E_m)}\langle m|\phi\big(|n\rangle\langle n|\big)|m\rangle$. The term $E_n - E_m$ is the decrease of energy in the reservoir, and by identifying this loss with the work performed, the analogy becomes evident. For the general transition between arbitrary matrix elements, the corresponding expression reads $\langle m|\mathcal{J}_{\beta H}\big(\phi\big(\mathcal{J}^{-1}_{\beta H}(|n\rangle\langle n'|)\big)\big)|m'\rangle = e^{\beta(E_n - E_m)/2} e^{\beta(E_{n'} - E_{m'})/2}\langle m|\phi\big(|n\rangle\langle n'|\big)|m'\rangle$. The fact that the off-diagonal case is governed by two energy differences, rather than one, corresponds to the evolution of the coherences in the energy reservoir.

7. Derivation of an intermediate fluctuation relation

Lemma 1. *With Assumptions 1, denote $P_i := |c_i\rangle\langle c_i|$ and $P_f := |c_f\rangle\langle c_f|$. Then it is the case that $V[e^{\alpha H^i_{S'}} \otimes P_i \otimes \hat{1}_E] = [e^{\alpha H^f_{S'}} \otimes P_f \otimes e^{\alpha H_E}]V[\hat{1}_{S'} \otimes P_i \otimes e^{-\alpha H_E}]$ and $[e^{\alpha H^f_{S'}} \otimes P_f \otimes \hat{1}_E]V = [\hat{1}_{S'} \otimes P_f \otimes e^{-\alpha H_E}]V[e^{\alpha H^i_{S'}} \otimes P_i \otimes e^{\alpha H_E}]$, for all $\alpha \in \mathbb{C}$.*

Proof. We only prove the first equality, since the proof of the second is analogous. First note that $e^{\alpha H^i_{S'}} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E = e^{\alpha H^i_{S'} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E}[\hat{1}_{S'} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E]$. Next we can use the fact that $H^i_{S'} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E = H - H^f_{S'} \otimes |c_f\rangle\langle c_f| \otimes \hat{1}_E - H^\perp \otimes \hat{1}_E - \hat{1}_{S'C} \otimes H_E$. Note that these summands commute with each other. Moreover, $H^f_{S'} \otimes |c_f\rangle\langle c_f| \otimes \hat{1}_E$ and $H^\perp \otimes \hat{1}_E$ have orthogonal supports compared to $\hat{1}_{S'} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E$. By these observations it follows that we can write

$$\begin{aligned} &V[e^{\alpha H^i_{S'}} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E] \\ &= V e^{\alpha H} e^{-\alpha \hat{1}_{S'C} \otimes H_E}[\hat{1}_{S'} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E] \\ &= e^{\alpha H} V[\hat{1}_{S'} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E] e^{-\alpha \hat{1}_{S'C} \otimes H_E}, \end{aligned}$$

where we in the second equality have used $[H, V] = 0$, and the fact that $\hat{1}_{S'} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E$ commutes with $e^{-\alpha \hat{1}_{S'C} \otimes H_E}$, as well as orthogonality of various terms. Next we use the assumed property of perfect control in Eq. (A1), i.e., $V[\hat{1}_{S'} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E] = [\hat{1}_{S'} \otimes |c_f\rangle\langle c_f| \otimes \hat{1}_E]V[\hat{1}_{S'} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E]$. When $e^{\alpha H}$ on the left hand side of V 'meets' $[\hat{1}_{S'} \otimes |c_f\rangle\langle c_f| \otimes \hat{1}_E]$, only the terms $e^{\alpha H^f_{S'} \otimes P^f_C \otimes \hat{1}_E} e^{\alpha \hat{1}_{S'C} \otimes H_E}$ survive. This leads to the first equality. The proof of the second equality is analogous. □

Proposition 1 (An intermediate quantum Crooks relation)**.** *With the definitions as in 1, the channels $\mathcal{F}$ and $\mathcal{R}$ defined in (A4) and (A5) are related as*

$$Z(H^i_{S'})\mathcal{F} = Z(H^f_{S'})\mathcal{J}_{\beta H_E}\mathcal{R}^*\mathcal{J}^{-1}_{\beta H_E}. \quad \text{(A9)}$$

With the separation of S' into system S and the heat bath B as in equation (A2) we thus get

$$Z(H_S^i)\mathcal{F} = Z(H_S^f)\mathcal{J}_{\beta H_E}\mathcal{R}^*\mathcal{J}_{\beta H_E}^{-1}. \tag{A10}$$

Proof. By comparing the definition (A5) of channel $\mathcal{R}$, with Eqs. (A6) and (A7) in section A 5 we can conclude that

$$\begin{aligned}
\mathcal{R}^*(Y) &= \mathrm{Tr}_{S'C}\big([G(H_{S'}^f)\otimes|c_f\rangle\langle c_f|\otimes\hat{1}_E] \\
&\qquad\times V[\hat{1}_{S'}\otimes\hat{1}_C\otimes Y]V^\dagger\big) \\
&= \frac{1}{Z(H_{S'}^f)}\mathrm{Tr}_{S'C}\Big([e^{-\beta H_{S'}^f/2}\otimes|c_f\rangle\langle c_f|\otimes\hat{1}_E]V \\
&\qquad\times[\hat{1}_{S'}\otimes\hat{1}_C\otimes Y] \\
&\qquad\times V^\dagger[e^{-\beta H_{S'}^f/2}\otimes|c_f\rangle\langle c_f|\otimes\hat{1}_E]\Big) \\
&\text{[By Lemma 1]} \\
&= \frac{1}{Z(H_{S'}^f)}e^{\beta H_E/2}\mathrm{Tr}_{S'C}\Big([\hat{1}_{S'}\otimes|c_f\rangle\langle c_f|\otimes\hat{1}_E]V \\
&\quad\times[e^{-\beta H_{S'}^i}\otimes|c_i\rangle\langle c_i|\otimes e^{-\beta H_E/2}Ye^{-\beta H_E/2}] \\
&\quad\times V^\dagger\Big)e^{\beta H_E/2} \\
&\text{[By perfect control, Eq. (A1)]} \\
&= \frac{Z(H_{S'}^i)}{Z(H_{S'}^f)}\mathcal{J}_{\beta H_E}^{-1}\circ\mathcal{F}\circ\mathcal{J}_{\beta H_E}(Y).
\end{aligned}$$

This can be rewritten as (A9).

With the additional assumption in equation (A2) we get $Z(H_{S'}^i) = Z(H_{S'}^i)Z(H_B)$ and $Z(H_{S'}^f) = Z(H_{S'}^f)Z(H_B)$. From this observation we obtain (A10) from (A9). □

Equation (A10) in Proposition 1 already has the flavor of a Crooks relation. However, as has been mentioned already, it relies on a too ambitious notion of reversal, where we invert the entire evolution. In the section C we shall remove this idealization, and assume that both the forward and reverse evolution is governed by the same 'direction of time', i.e., in both cases the global evolution is given by applying the map $Q \mapsto VQV^\dagger$.

8. Remarks concerning Crooks operation time reversal and Petz recovery channel

The notion of 'operation time reversals' was introduced in [46] as a quantum generalization of time-reversals of classical Markov chains. Given a channel Φ with fix-point density operator ρ, i.e., $\Phi(\rho) = \rho$, the operation time reversal of Φ is defined by the mapping $\sigma \mapsto \sqrt{\rho}\Phi^*(\sqrt{\rho}^{-1}\sigma\sqrt{\rho}^{-1})\sqrt{\rho}$. Let us now compare this with the right hand side of (A9). By construction $\mathcal{R}$ is a channel. However, one can confirm that it is the case that

$$\mathcal{R}(e^{-\beta H_E}) = \frac{Z(H_{S'}^i)}{Z(H_{S'}^f)}e^{-\beta H_E}. \tag{A11}$$

Hence, $e^{-\beta H_E}$ is *not* a fix point of $\mathcal{R}$, and thus the conditions for Crooks time reversal are not quite satisfied (unless $Z(H_{S'}^i) = Z(H_{S'}^f)$, which would be the case for a cyclic process, i.e., if $H_{S'}^f = H_{S'}^i$).

There exists a more general construction introduced in information theory, namely Petz recovery channel [47–50]. Given a channel Φ and a reference state ρ (that does not have to be a fix point of the channel Φ) Petz recovery channel is defined as

$$\widehat{\Phi}(\sigma) = \sqrt{\rho}\Phi^*\big(\Phi(\rho)^{-1/2}\sigma\Phi(\rho)^{-1/2}\big)\sqrt{\rho}. \tag{A12}$$

Hence, Crooks time reversal emerges as a special case when ρ is a fix point of Φ.

If we take $e^{-\beta H_E}$ as the reference operator, it is straightforward to confirm that the intermediate fluctuation relation (A9) can be rephrased as $\mathcal{F} = \widehat{\mathcal{R}}$, i.e., that the forward channel is equal to the Petz-transformation (A12) of the reverse channel.

In [40] it was shown that the work gain in work extraction can be bounded by how well the initial state can be reconstructed via Petz recovery channel. One can in particular note the similarity between our channel $\mathcal{R}$ and the recovery channel $\mathcal{R}_{\rho\to\sigma}$ defined in [40], although the latter is a proper thermal operation, while the definition of $\mathcal{R}$ in addition contains the control system that is subject to perfect control. In view of these structural similarities it is tempting to speculate on deeper relations between these results. Fluctuation relations can be viewed as statistical manifestations of the second law [6]; an observation that makes the connection to work extraction more plausible. Moreover, there are investigations that hint or elaborate on connections. One example is the generalized Jarzynski relations for feedback control [103, 118, 119, 139]. Moreover, in [31] a classical Crook's relation was used as a component in a proof about single-shot work extraction, and recent investigations [133–136] have focused on exploring links between single-shot concepts and fluctuation theorems. On a similar note one may wonder whether there exist a more operational characterization of fluctuation relations. Although intriguing questions, we will not consider them further in this investigation.

Appendix B: Time reversal and time reversal symmetry

Here we introduce the notion of time-reversals that we use to define our quantum Crooks relation. As mentioned in the main text, our time-reversals are primarily related to transposes rather than to complex conjugates (cf. the discussions on time-reversals in [56, 57]). Before we proceed with the definition we do in the following subsection compare complex conjugation and transposes regarded as time-reversals.

1. Complex conjugation vs transpose

In standard textbooks on quantum mechanics (see e.g. section 4.4 in [54]) time reversal is often expressed on the level of Hilbert spaces as complex conjugation $\psi^*(x)$ of wave functions $\psi(x)$, or via an orthonormal basis as $|\psi^*\rangle = \sum_n |n\rangle\langle n|\psi\rangle^*$. On the level of operators, this translates to $Q^* = \int\int \langle x|Q|x'\rangle^* |x\rangle\langle x'|dxdx'$, or $Q^* = \sum_{nn'} \langle n|Q|n'\rangle^* |n\rangle\langle n'|$. In comparison, the transpose acts as $Q^* = \int\int |x\rangle\langle x'|Q|x\rangle\langle x'|dxdx'$, or $Q^t = \sum_{nn'} |n\rangle\langle n'|Q|n\rangle\langle n'|$.

Both the complex conjugate and the transpose implements time-reversals, but in a sightly different manner. Suppose that a Hermitian generator for the time evolution H_{evol} satisfies $H^*_{\text{evol}} = H_{\text{evol}}$, or equivalently $H^t_{\text{evol}} = H_{\text{evol}}$. (One should not confuse H_{evol}, discussed in Sec. A 2, with the Hamiltonians H, H_S, H_E etc.) The time-evolution operator transforms as

$$\begin{aligned}(e^{-itH_{\text{evol}}/\hbar})^* =& e^{itH_{\text{evol}}/\hbar},\\ (e^{-itH_{\text{evol}}/\hbar})^t =& e^{-itH_{\text{evol}}/\hbar}.\end{aligned} \tag{B1}$$

Hence, complex conjugation inverts the evolution operator, while the transpose leaves it intact. At first sight it might thus seem a bit odd that the transpose can implement any form of time-reversal. To understand this we should consider the manner in which these mappings act on products of operators, namely

$$(AB)^* = A^*B^*, \quad (AB)^t = (B)^t(A)^t. \tag{B2}$$

In other words, complex conjugation leaves the operator ordering intact, while the transpose reverses the ordering. In some sense, (B1) and (B2) complement each other when it comes to the reversal of the evolution. To see this, assume that an initial state ρ_i is evolved into the state $\rho_f = e^{-itH_{\text{evol}}/\hbar}\rho_i e^{itH_{\text{evol}}/\hbar}$. For the complex conjugate we get

$$\begin{aligned}\rho_f^* =& (e^{-itH_{\text{evol}}/\hbar})^* \rho_i^* (e^{itH_{\text{evol}}/\hbar})^*\\ =& e^{itH_{\text{evol}}/\hbar}\rho_i^* e^{-itH_{\text{evol}}/\hbar},\end{aligned}$$

and hence $\rho_i^* = e^{-itH_{\text{evol}}/\hbar}\rho_f^* e^{itH_{\text{evol}}/\hbar}$. Similarly,

$$\begin{aligned}\rho_f^t =& (e^{itH_{\text{evol}}/\hbar})^t \rho_i^t (e^{-itH_{\text{evol}}/\hbar})^t\\ =& e^{itH_{\text{evol}}/\hbar}\rho_i^t e^{-itH_{\text{evol}}/\hbar},\end{aligned}$$

and consequently $\rho_i^t = e^{-itH_{\text{evol}}/\hbar}\rho_f^t e^{itH_{\text{evol}}/\hbar}$. Hence, we do again obtain the effective reversal of the evolution. One can see that for the complex conjugation the reversal this is due to to the fact that the complex conjugation inverts the time evolution operator, while for the transpose it is due to the inversion of the operator ordering.

2. What we require from time-reversals

Instead of directly defining time-reversals in terms of transposes we here rather define it via a 'wish-list' of properties. By inspection one can see that transposes satisfy these conditions, although the latter allow for a slightly larger class of operations (see Proposition 2). One can also see that this definition immediately excludes the complex conjugation (due to the assumed linearity). Hence, one should not take this list as the ultimate and most general definition of what a time-reversal possibly could be, but rather as convenient set of assumptions that is sufficient for our purposes and makes the book-keeping in the proofs simple. It may potentially be the case that a more general notion of time-reversals could extend the resulting family of quantum Crooks relations. Although an interesting question, it will not be pursued further in this investigation.

Definition 1. *A linear map $\mathcal{T}$ is called a time-reversal if*

$$\mathcal{T}(AB) = \mathcal{T}(B)\mathcal{T}(A), \tag{B3a}$$

$$\mathcal{T}(A^\dagger) = \mathcal{T}(A)^\dagger, \tag{B3b}$$

$$\text{Tr}[\mathcal{T}(\sigma)] = \text{Tr}(\sigma), \tag{B3c}$$

$$\mathcal{T}^2 = I. \tag{B3d}$$

As a bit of a technical remark, in the infinite-dimensional case one may additionally require that $\mathcal{T}$ maps bounded operators to bounded operators, and trace class operators to trace class operators. If one restricts to bounded A, B in (B3a) it follows that $\mathcal{T}(A)$, $\mathcal{T}(B)$, and $\mathcal{T}(AB)$ are bounded. By demanding that A in (B3b) is bounded we make sure that the Hilbert adjoint $A^\dagger$ is well defined and bounded (see Theorem 3.9-2 in [140]). By the requirement that $\mathcal{T}$ maps bounded operators to bounded operators we know that $\mathcal{T}(A)$ is bounded, and thus $\mathcal{T}(A)^\dagger$ is also well defined. If one restricts σ to be trace class in (B3c) it follows that $\text{Tr}(\sigma)$ is well defined, and if $\mathcal{T}$ maps trace class operators to trace class operators, $\text{Tr}[\mathcal{T}(\sigma)]$ is also well defined. Although this is a reasonable collection of assumptions, one should keep in mind that we here tend to apply these maps also to unbounded operators.

3. The $\ominus$-transformation

For a CPM ϕ we define $\phi^\ominus$ as

$$\phi^\ominus := \mathcal{T}\phi^*\mathcal{T}, \tag{B4}$$

where $\mathcal{T}$ is a given time-reversal, and where ϕ^* is the conjugation, as discussed in section A 5. It is a straightforward application of the properties of the time-reversal $\mathcal{T}$ to show the following alternative definition

$$\phi^\ominus = (\mathcal{T}\phi\mathcal{T})^*. \tag{B5}$$

It is also straightforward to confirm the following lemma.

Lemma 2. *If ϕ is a CPM with Kraus decomposition $\phi(\sigma) = \sum_k V_k \sigma V_k^\dagger$, and $\mathcal{T}$ is a time-reversal, then*

$$\phi^\ominus(\sigma) = \sum_k \mathcal{T}(V_k)\sigma\mathcal{T}(V_k)^\dagger. \tag{B6}$$

In other words, if $\{V_k\}_k$ is a Kraus representation of ϕ, then $\{\mathcal{T}(V_k)\}_k$ is a Kraus representation of $\phi^\ominus$. Hence, if ϕ is a CPM, then $\phi^\ominus$ is a CPM. If ϕ is a channel (trace preserving CPM), then $\phi^\ominus$ is not necessarily a channel. However, if ϕ is a unital channel ($\phi(\hat{1}) = \hat{1}$), then $\phi^\ominus$ is a channel, and moreover a unital channel.

4. Characterization of $\mathcal{T}$ in finite dimensions

The purpose of this section is to make more precise what kind of mappings that the list of properties in Definition 1 specifies, how they relate to transposes, as well as deriving some further properties that will be useful for the subsequent derivations. Throughout this section we assume that the underlying Hilbert space is finite-dimensional, although some of the results would be straightforward to extend to the infinite-dimensional case.

For a finite-dimensional Hilbert space $\mathcal{H}$, we do in the following let $\mathcal{L}(\mathcal{H})$ denote the set of linear operators on $\mathcal{H}$. Our first general observation is that if $\mathcal{T}_1$ and $\mathcal{T}_2$ are time-reversals on two different finite-dimensional Hilbert spaces, then $\mathcal{T}_1 \otimes \mathcal{T}_2$ is also a time-reversal. It turns out that each single property in Definition 1 is separately preserved under the tensor product. The proof can be obtained via decompositions $Q = \sum_{mn} Q_1^{(m)} \otimes Q_2^{(n)}$, where $Q_1^{(m)}$ and $Q_2^{(n)}$ are operators on $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively.

Lemma 3. *If $\mathcal{T}$ is a linear map that satisfies conditions (B3a) and (B3d), then $\mathcal{T}(\hat{1}) = \hat{1}$.*

Proof. By applying $\mathcal{T}$ to the trivial identity $\mathcal{T}(\hat{1}) = \mathcal{T}(\hat{1})\hat{1}$ and use property (B3d) it follows that $\hat{1} = \mathcal{T}\big(\mathcal{T}(\hat{1})\hat{1}\big) = \mathcal{T}(\hat{1})\mathcal{T}(\mathcal{T}(\hat{1})) = \mathcal{T}(\hat{1})\hat{1} = \mathcal{T}(\hat{1})$, where the second equality follows by (B3a). □

Lemma 4. *Let $\mathcal{T}$ be linear. If $\mathcal{T}$ satisfies (B3a) and (B3b), then $\mathcal{T}$ is positive, i.e., $Q \geq 0 \Rightarrow \mathcal{T}(Q) \geq 0$.*

Proof. If $Q \geq 0$ then there exists A such that $Q = AA^\dagger$. Hence $\mathcal{T}(Q) = \mathcal{T}(AA^\dagger) = \mathcal{T}(A^\dagger)\mathcal{T}(A) = \mathcal{T}(A)^\dagger\mathcal{T}(A)$, where the second equality follows by (B3a), and the third by (B3b). Hence, $\mathcal{T}(Q) \geq 0$. □

Lemma 5. *Let $\mathcal{T}$ be a linear map.*

- *If $\mathcal{T}$ satisfies (B3a) and (B3b), then $\mathcal{T}$ maps projectors to projectors. Furthermore, pairwise orthogonal projectors get mapped to pairwise orthogonal projectors.*
- *If $\mathcal{T}$ satisfies (B3a), (B3b), and (B3c), then $\mathcal{T}$ preserves the dimension of the projected subspaces. In particular $\mathcal{T}$ preserves purity, i.e. if $|\psi\rangle \in \mathcal{H}$ is normalized, then there exists a normalized $|\chi_\psi\rangle \in \mathcal{H}$ such that $\mathcal{T}(|\psi\rangle\langle\psi|) = |\chi_\psi\rangle\langle\chi_\psi|$.*
- *If $\mathcal{T}$ satisfies (B3a), (B3b), (B3c), and if the underlying Hilbert space is finite-dimensional, then $\mathcal{T}(\hat{1}) = \hat{1}$.*

For the third item it is necessary to restrict to finite dimensions. As an example, let $\{|n\rangle\}_{n\in\mathbb{N}}$ be a complete orthonormal basis, and let $\mathcal{T}(|n\rangle\langle n'|) = |2n'\rangle\langle 2n|$. This satisfies (B3a), (B3b) and (B3c), but $\mathcal{T}(\hat{1}) \neq \hat{1}$.

Proof. A linear operator P is a projector if and only if $P^2 = P$ and $P^\dagger = P$ (and P is bounded in the infinite-dimensional case). Assuming that P is a projector it follow by properties (B3a) and (B3b) that $\mathcal{T}(P)$ is a also a projectors. (If $\mathcal{T}$ preserves boundedness, then the boundedness of $\mathcal{T}(P)$ is guaranteed.) Two projectors are orthogonal if and only if $P_1P_2 = 0$. Thus by property (B3a) it follows that $\mathcal{T}(P_2)\mathcal{T}(P_1) = \mathcal{T}(P_1P_2) = 0$. The dimension of the subspace onto which a projector P projects is given by $\mathrm{Tr}(P)$. By assumption (B3c) it follows that $\mathrm{Tr}(\mathcal{T}(P)) = \mathrm{Tr}(P)$. Hence, the dimension is preserved. In the case that the Hilbert space is finite-dimensional, then $\mathrm{Tr}\mathcal{T}(\hat{1}) = \mathrm{Tr}\hat{1}$ is the dimension of the Hilbert space. Hence $\mathcal{T}(\hat{1})$ is a projector with the dimension of the Hilbert space, and thus $\mathcal{T}(\hat{1}) = \hat{1}$. □

In the following we denote the standard operator norm by $\|Q\| := \sup_{\psi:|\psi\|=1} \|Q|\psi\rangle\|$, and the trace norm by $\|Q\|_1 := \mathrm{Tr}\sqrt{QQ^\dagger} = \mathrm{Tr}\sqrt{Q^\dagger Q}$ (where the last equality follows by the singular value decomposition of Q in the finite-dimensional case).

Lemma 6. *Let $\mathcal{T}$ be a time-reversal as in Definition 1, then $\|\mathcal{T}(Q)\| = \|Q\|$, $\|\mathcal{T}(Q)\|_1 = \|Q\|_1$.*

Proof. First we note that $\|\mathcal{T}(Q)|\psi\rangle\|^2 = \mathrm{Tr}\big(|\psi\rangle\langle\psi|\mathcal{T}(QQ^\dagger)\big) = \mathrm{Tr}\big(QQ^\dagger|\chi_\psi\rangle\langle\chi_\psi|\big) = \|Q^\dagger|\chi_\psi\rangle\|^2$, where $|\chi_\psi\rangle$ is such that $|\chi_\psi\rangle\langle\chi_\psi| = \mathcal{T}(|\psi\rangle\langle\psi|)$ as in Lemma 5. Consequently, $\|\mathcal{T}(Q)\| \leq \|Q^\dagger\|$. By $\|Q^\dagger\| = \|Q\|$ (see, e.g., Theorem 3.9-2 in [140]) it thus follows that $\|\mathcal{T}(Q)\| \leq \|Q\|$. By substituting Q with $\mathcal{T}(Q)$ in the above reasoning, and use $\mathcal{T}^2 = I$ one obtains $\|Q\| \leq \|\mathcal{T}(Q)\|$. Hence, $\|\mathcal{T}(Q)\| = \|Q\|$.

Next we make the observation that $\|\mathcal{T}(Q)\|_1 = \mathrm{Tr}\sqrt{\mathcal{T}(Q)\mathcal{T}(Q)^\dagger} = \mathrm{Tr}\sqrt{\mathcal{T}(Q^\dagger Q)}$. By Lemma 4 we know that $\mathcal{T}$ maps positive operators to positive operators. Hence, $\mathcal{T}(Q^\dagger Q)$ is a positive operator, and thus $\sqrt{\mathcal{T}(Q^\dagger Q)}$ is well defined and positive (see e.g. section 9.4 in [140]). By Lemma 4 we know that $\mathcal{T}(\sqrt{Q^\dagger Q}) \geq 0$. Moreover, $\mathcal{T}(\sqrt{Q^\dagger Q})\mathcal{T}(\sqrt{Q^\dagger Q}) = \mathcal{T}(Q^\dagger Q)$. By the reasoning above we thus know that both $\sqrt{\mathcal{T}(Q^\dagger Q)}$ and $\mathcal{T}(\sqrt{Q^\dagger Q})$ are positive square roots of $\mathcal{T}(Q^\dagger Q)$. However, the positive square root of a positive operator is unique (Proposition 9.4-2 in [140]), and thus

$\sqrt{\mathcal{T}(Q^\dagger Q)} = \mathcal{T}(\sqrt{Q^\dagger Q})$. Consequently $\|\mathcal{T}(Q)\|_1 = \mathrm{Tr}\sqrt{\mathcal{T}(Q^\dagger Q)} = \mathrm{Tr}\mathcal{T}(\sqrt{Q^\dagger Q}) = \|Q\|_1$. □

Lemma 7. *On a finite-dimensional complex Hilbert space, let H be Hermitian with an orthogonal family of eigenprojectors $\{P_m\}_m$ and corresponding eigenvalues h_m, such that $h_m \neq h_{m'}$ whenever $m \neq m'$, and $H = \sum_m h_m P_m$. If $\mathcal{T}$ is a time reversal such that $\mathcal{T}(H) = H$, then $\mathcal{T}(P_m) = P_m$. Hence, $\mathcal{T}$ preserves the eigenspaces.*

Proof. By Lemma 5 we know that each $\mathcal{T}(P_m)$ is a projector, and that it projects on a subspace of the same dimension as P_m. Next one can confirm that $H\mathcal{T}(P_m) = \mathcal{T}(H)\mathcal{T}(P_m) = \mathcal{T}(P_m H) = h_m \mathcal{T}(P_m)$. Hence, $\mathcal{T}(P_m)$ must be an eigenprojector corresponding to eigenvalue h_m. Since $\mathcal{T}(P_m)$ and P_m projects on spaces of the same dimension, we must have $\mathcal{T}(P_m) = P_m$. □

Given an orthonormal basis $\{|k\rangle\}_{k=1}^K$ of a finite-dimensional Hilbert space $\mathcal{H}$, we define the transpose with respect to this basis as

$$Q^t := \sum_{kk'} |k'\rangle\langle k|Q|k'\rangle\langle k|. \tag{B7}$$

Since the transpose is dependent of the choice of basis, an obvious question is what happens when we make a change of basis. The following lemma, which we state without proof (it is straightforward to verify) specifies how one can express the new transpose in terms of the old.

Lemma 8. *On a finite-dimensional complex Hilbert space $\mathcal{H}$ let the transpose t_{old} be defined with respect to an orthonormal basis $\{|\mathrm{old}_k\rangle\}_k$. Let the transpose t_{new} be defined with respect to the orthonormal basis $\{|\mathrm{new}_k\rangle\}_k$, where $|\mathrm{new}_k\rangle := W|\mathrm{old}_k\rangle$ for some unitary operator W on $\mathcal{H}$. Then $W^{t_{new}} = WW^{t_{\mathrm{old}}}W^\dagger$, and the new transpose t_{new} can be expressed in terms of the old basis as*

$$Q^{t_{\mathrm{new}}} = WW^{t_{\mathrm{old}}}Q^{t_{\mathrm{old}}}(WW^{t_{\mathrm{old}}})^\dagger. \tag{B8}$$

Similarly, the old transpose t_{old} can be expressed in terms of the new basis as

$$Q^{t_{old}} = (W^{t_{new}}W)^\dagger Q^{t_{new}} W^{t_{new}} W. \tag{B9}$$

The following Lemma is a special case of Autonne-Takagi's decomposition, see e.g. Corollary 4.4.4 in [141].

Lemma 9 (Special case of Autonne-Takagi's decomposition)**.** *Let $\mathcal{H}$ be finite-dimensional complex Hilbert space. Let U be a unitary operator on $\mathcal{H}$. Then $U^t = U$ (with respect to a given orthonormal basis of $\mathcal{H}$) if and only if there exists a unitary operator W on $\mathcal{H}$ such that $U = WW^t$.*

By combining Lemma 9 with Lemma 8 we can conclude that transformations of transposes are characterized by complex symmetric unitary operators.

Proposition 2. *Let $\mathcal{H}$ be a finite-dimensional complex Hilbert space, and let $\mathcal{B} := \{|k\rangle\}_{k=1}^K$ be an orthonormal basis of $\mathcal{H}$. Let t denote the transpose with respect to the basis $\mathcal{B}$. Let $\mathcal{T}$ be a linear map on $\mathcal{L}(\mathcal{H})$.*

1. *$\mathcal{T}$ satisfies (B3a), (B3b), and (B3c) if and only if there exists a unitary operator U on $\mathcal{H}$ such that*

$$\mathcal{T}(|k\rangle\langle k'|) = U|k'\rangle\langle k|U^\dagger, \quad \forall k, k'. \tag{B10}$$

or equivalently

$$\mathcal{T}(Q) = UQ^tU^\dagger, \quad \forall Q \in \mathcal{L}(\mathcal{H}). \tag{B11}$$

Moreover, U is uniquely determined by $\mathcal{T}$ and $\{|k\rangle\}_k$ up to a global phase factor.

2. *If $\mathcal{T}$ satisfies (B3a), (B3b), and (B3c), then the following are equivalent:*
   - *$\mathcal{T}$ satisfies (B3d).*
   - *The unitary operator U in (B10) satisfies $U^t = \pm U$, i.e., U is complex symmetric or complex skew-symmetric. (The choice of global phase factor in U does not affect the property of being symmetric or skew-symmetric.)*

3. *If $\mathcal{T}$ satisfies (B3a), (B3b), and (B3c), then the following are equivalent:*
   - *There exists an orthonormal basis $\{|\xi_k\rangle\}_k$ of $\mathcal{H}$ such that*

$$\mathcal{T}(|\xi_k\rangle\langle\xi_{k'}|) = |\xi_{k'}\rangle\langle\xi_k|, \quad \forall k, k'. \tag{B12}$$

   - *The unitary operator U in (B10) satisfies $U^t = U$, i.e., if U is complex symmetric.*

As a further remark one may note that (B11) directly implies that $\mathcal{T}$ is a positive but not completely positive map, since it is a composition of a unitary operation and a transpose, and the transpose is not completely positive [58, 59].

Proof of Proposition 2. We start by proving that properties (B3a), (B3b), (B3c) implies equation (B10). From Lemma 5 we know that $\{\mathcal{T}(|k\rangle\langle k|)\}_k$ is a set of pairwise orthogonal projectors onto one-dimensional subspaces. This means that there exists a unitary operator $\tilde{U}$ such that $\mathcal{T}(|k\rangle\langle k|) = \tilde{U}|k\rangle\langle k|\tilde{U}^\dagger$. Moreover, by (B3a) it follows that $\mathcal{T}(|k\rangle\langle k'|) = \mathcal{T}(|k\rangle\langle k||k\rangle\langle k'||k'\rangle\langle k'|) = z_{kk'}\tilde{U}|k'\rangle\langle k|\tilde{U}^\dagger$, where $z_{kk'} := \langle k'|\tilde{U}^\dagger\mathcal{T}(|k\rangle\langle k'|)\tilde{U}|k\rangle$. By (B3b) it follows that $z^*_{kk'} = z_{k'k}$ By using $\mathcal{T}(|k\rangle\langle k|) = \tilde{U}|k\rangle\langle k|\tilde{U}^\dagger$ and (B3a) it follows that $z_{k'k}z_{kk''} = \langle k''|\tilde{U}^\dagger\mathcal{T}(|k\rangle\langle k''|)\mathcal{T}(|k\rangle\langle k|)\mathcal{T}(|k'\rangle\langle k|)\tilde{U}|k'\rangle = z_{k'k''}$. One can realize that these two last conditions together imply that $z_{k'k''} = z_{1k'}z^*_{1k''}$, and next that $|z_{1k}| = 1$. Hence, there exist real numbers θ_k such that $z_{k'k''} = e^{i(\theta_{k'}-\theta_{k''})}$. By putting $U := \tilde{U}\sum_k e^{-i\theta_k}|k\rangle\langle k|$, we find that (B10) holds. For the opposite implication,

assume that there exists a unitary U such that (B11) holds. It is straightforward to confirm that each of the properties (B3a), (B3b), (B3c) are satisfied.

For uniqueness, suppose that that there exist two unitary operators U_1, U_2 that both satisfy (B11). Consequently, $U_1 Q^t U_1^\dagger = U_2 Q^t U_2^\dagger$ for all $Q \in \mathcal{L}(\mathcal{H})$, from which it follows that $U_1 = e^{i\chi} U_2$ for some $\chi \in \mathbb{R}$

Next, we turn to the second item of the proposition. Assume that (B3a), (B3b), and (B3c) are satisfied. We know that there exists a unitary operator U such that equation (B11) holds. If we use this observation twice we find

$$\begin{aligned}\mathcal{T}(\mathcal{T}(Q)) =& \mathcal{T}(UQ^tU^\dagger) = U(UQ^tU^\dagger)^tU^\dagger \\ =& (U^tU^\dagger)^\dagger Q U^t U^\dagger.\end{aligned} \tag{B13}$$

This implies that $\mathcal{T}^2 = I$ if and only if $U^t = e^{i\theta}U$ for some real number θ. By the definition of the transpose it follows that $\langle k'|U|k\rangle = e^{i\theta}\langle k|U|k'\rangle$ for all k, k'. If this equality is iterated we obtain $\langle k'|U|k\rangle = e^{i\theta}\langle k|U|k'\rangle = e^{2i\theta}\langle k'|U|k\rangle$, and thus $(1-e^{2i\theta})\langle k'|U|k\rangle = 0$, for all k, k'. Hence, either $1 - e^{2i\theta} = 0$, or $\langle k'|U|k\rangle = 0$ for all k, k'. However, the latter is not possible since U is unitary. We can conclude that $e^{2i\theta} = 1$, and thus $e^{i\theta} = \pm 1$. This combined with $U^t = e^{i\theta}U$ yields $U^t = \pm U$.

Finally we turn to the third item of the proposition. Let t' denote the transpose with respect to $\{|\xi_k\rangle\}_k$. Then (B12) is the same as saying that $\mathcal{T}(Q) = Q^{t'}$. By Lemma 8 we know that we can express the transpose t' in terms of the transpose t (with respect to the basis $\{|k\rangle\}_k$) as $Q^{t'} = WW^tQ^t(WW^t)^\dagger$, for a unitary operator W such that $|\xi_k\rangle = W|k\rangle$. In terms of the original basis $\{|k\rangle\}_k$ we know that $\mathcal{T}(Q) = UQ^tU^\dagger$. Hence, $UQ^tU^\dagger = WW^tQ^t(WW^t)^\dagger$, and thus $U = WW^te^{i\chi}$ for some $\chi \in \mathbb{R}$. Hence, we can conclude that the unitary operator U satisfies $U^t = U$ and thus is complex symmetric.

To derive the opposite implication, assume that there exists a unitary operator U such that $U^t = U$ and (B10) holds. By Lemma 9 we know that there exist a unitary operator W on $\mathcal{H}$ such that $U = WW^t$. Define $|\xi_l\rangle := W|l\rangle$ for all l. Since W is unitary it follows that $\{|\xi_l\rangle\}_l$ is an orthonormal basis of $\mathcal{H}$. One can verify that (B12) holds. □

Appendix C: A quantum fluctuation theorem

1. Re-setting the stage

Here we construct a new set of assumptions that includes time reversal symmetry (see Fig. 9).

Assumptions 2. *Let $\mathcal{H}_{S'}$, $\mathcal{H}_C$, and $\mathcal{H}_E$ be complex Hilbert spaces. Let $|c_{i+}\rangle, |c_{i-}\rangle, |c_{f+}\rangle, |c_{f-}\rangle \in \mathcal{H}_C$ be normalized and such that the linear span $\mathcal{H}_C^i := \mathrm{Sp}\{|c_{i+}\rangle, |c_{i-}\rangle\}$ is orthogonal to $\mathcal{H}_C^f := \mathrm{Sp}\{|c_{f+}\rangle, |c_{f-}\rangle\}$. Let P_C^i and P_C^f denote the projectors onto these two subspaces, and define $P_C^\perp := \hat{1}_C - P_C^i - P_C^f$.*

- *$H_{S'}^i$ and $H_{S'}^f$ are Hermitian operators on $\mathcal{H}_{S'}$ such that $Z(H_{S'}^i)$ and $Z(H_{S'}^f)$ are finite. Let H_E be an Hermitian operator on $\mathcal{H}_E$. Let $H^\perp$ be a Hermitian operator on $\mathcal{H}_{S'} \otimes \mathcal{H}_C$ such that $[\hat{1}_{S'} \otimes P_C^\perp]H^\perp[\hat{1}_{S'} \otimes P_C^\perp] = H^\perp$, and define*

$$H_{S'C} := H_{S'}^i \otimes P_C^i + H_{S'}^f \otimes P_C^f + H^\perp.$$

and $H := H_{S'C} \otimes \hat{1}_E + \hat{1}_{S'C} \otimes H_E$.

- *$\mathcal{T}_{S'C}$ and $\mathcal{T}_E$ are time-reversals, and $\mathcal{T} := \mathcal{T}_{S'C} \otimes \mathcal{T}_E$. We assume*

$$\mathcal{T}_{S'C}(H_{S'C}) = H_{S'C}, \quad \mathcal{T}_E(H_E) = H_E, \tag{C1}$$

and

$$\begin{aligned}\mathcal{T}_{S'C}(\hat{1}_{S'} \otimes |c_{i+}\rangle\langle c_{i+}|) =& \hat{1}_{S'} \otimes |c_{i-}\rangle\langle c_{i-}|, \\ \mathcal{T}_{S'C}(\hat{1}_{S'} \otimes |c_{f+}\rangle\langle c_{f+}|) =& \hat{1}_{S'} \otimes |c_{f-}\rangle\langle c_{f-}|.\end{aligned} \tag{C2}$$

- *V is unitary operator on $\mathcal{H}_{S'} \otimes \mathcal{H}_C \otimes \mathcal{H}_E$ such that $[V, H] = 0$, $\mathcal{T}(V) = V$, and*

$$\begin{aligned}&V[\hat{1}_{S'} \otimes |c_{i+}\rangle\langle c_{i+}| \otimes \hat{1}_E] \\ &= [\hat{1}_{S'} \otimes |c_{f+}\rangle\langle c_{f+}| \otimes \hat{1}_E]V.\end{aligned} \tag{C3}$$

In sections G 3 c and G 3 d it is demonstrated that there exist setups that satisfy all conditions in Assumptions 2.

These assumptions are constructed in such a way that the unitary V, the initial state $|c_i\rangle := |c_{i+}\rangle$, and the final state $|c_f\rangle := |c_{f+}\rangle$ satisfy Assumptions 1. This translation only mounts to redefining the projector $P_C^\perp$ and the Hamiltonian $H^\perp$. More precisely, starting with Assumptions 2 we can define the projectors $P_{ci+} := P_C^i - |c_{i+}\rangle\langle c_{i+}|$ and $P_{cf+} := P_C^f - |c_{f+}\rangle\langle c_{f+}|$. With $P_C^\perp$ being the projector in Assumptions 2, we can define the new projector in Assumptions 1 as $\overline{P}_C^\perp := P_C^\perp + P_{ci+} + P_{cf+}$. Similarly, given the Hamiltonian $H^\perp$ in Assumptions 2 we can define the new $\overline{H}^\perp := H_{S'}^i \otimes P_{ci+} + H_{S'}^f \otimes P_{cf+} + H^\perp$ in Assumptions 1. All other Hamiltonians and operators can be left unchanged. Hence, the restriction to Assumptions 1 only requires us to reshuffle the Hamiltonians. In an analogous manner, V, $|c_i\rangle := |c_{f-}\rangle$, and $|c_f\rangle := |c_{i-}\rangle$ also form a valid triple in Assumptions 1. Note that in this case $|c_{f-}\rangle$ is the initial state of the effectively reversed evolution.

One should note that one can consider several variations on Assumptions 2. One could for example imagine an alternative to (C2) where we instead assume a product time-reversal $\mathcal{T}_{S'C} = \mathcal{T}_{S'} \otimes \mathcal{T}_C$, and demand $\mathcal{T}_C(|c_{i+}\rangle\langle c_{i+}|) = |c_{i-}\rangle\langle c_{i-}|$, $\mathcal{T}_C(|c_{f+}\rangle\langle c_{f+}|) = |c_{f-}\rangle\langle c_{f-}|$. However, the assumption in (C2) is more general, and provides a rather useful flexibility.

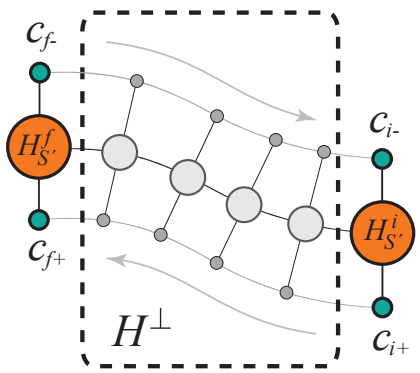

FIG. 9: **Structure of the Hamiltonian.** The Hamiltonian $S'C$ is of the form $H_{S'C} = H^i_{S'} \otimes P^i_C + H^f_{S'} \otimes P^f_C + H^\perp$. Hence, whether the state of the control is in subspace $\mathcal{H}^i_C$ onto which P^i_C projects, or in $\mathcal{H}^f_C$ onto which P^f_C projects, determines the Hamiltonian of S'. The Hamiltonian $H^\perp$ corresponds to any possible intermediate stages. The initial control space $\mathcal{H}^i_C$ is spanned by two states $|c_{i+}\rangle$ and $|c_{i-}\rangle$, which are the time-reversals of each other. Analogously, the final control space $\mathcal{H}^f_C$ is spanned by $|c_{f+}\rangle$ and $|c_{f-}\rangle$. The global evolution V is such that it brings control state $|c_{i+}\rangle$ into $|c_{f+}\rangle$, while it brings $|c_{f-}\rangle$ into $|c_{i-}\rangle$, thus implementing both the forward and the reverse process.

One should keep in mind that we do not assume that $|c_{i+}\rangle$ necessarily is orthogonal to $|c_{i-}\rangle$. (However, we do assume that $\mathrm{Sp}\{|c_{i+}\rangle, |c_{i-}\rangle\}$ is orthogonal to $\mathrm{Sp}\{|c_{f+}\rangle, |c_{f-}\rangle\}$.)

Due to the time-reversal symmetry, a perfect transition from $|c_{i+}\rangle$ to $|c_{f+}\rangle$ implies a perfect transition of $|c_{f-}\rangle$ to $|c_{i-}\rangle$. More precisely, by combining (C3) with the properties $\mathcal{T}(AB) = \mathcal{T}(B)\mathcal{T}(A)$, $\mathcal{T}_E(\hat{1}_E) = \hat{1}_E$, as well as the assumptions $\mathcal{T}(V) = V$ and (C2), one obtains

$$\begin{aligned} &V[\hat{1}_{S'} \otimes |c_{f-}\rangle\langle c_{f-}| \otimes \hat{1}_E] \\ &= [\hat{1}_{S'} \otimes |c_{i-}\rangle\langle c_{i-}| \otimes \hat{1}_E]V. \end{aligned} \tag{C4}$$

As the reader may have noticed, a considerable part of Assumptions 2 deals with the control system, which is due to the rather strong idealization that perfect control entails. In section H we will abandon this idealization, and as a bonus we also obtain a more lean set of assumptions (cf. Assumptions 4).

Lemma 10. *With Assumptions 2 it is the case that*

$$\begin{aligned} \mathcal{T}_{S'C}\big(G(H^i_{S'}) \otimes |c_{i\pm}\rangle\langle c_{i\pm}|\big) =& G(H^i_{S'}) \otimes |c_{i\mp}\rangle\langle c_{i\mp}|, \\ \mathcal{T}_{S'C}\big(G(H^f_{S'}) \otimes |c_{f\pm}\rangle\langle c_{f\pm}|\big) =& G(H^f_{S'}) \otimes |c_{f\mp}\rangle\langle c_{f\mp}|. \end{aligned} \tag{C5}$$

One may wonder why we do not directly assume (C5) in Assumptions 2 rather than (C2). The reason is partially that the latter choice defines the action of the time reversal on the control states in a more clean manner, but also because it aligns with the more general set of assumptions that we will use in section H 3.

Proof. We first note that $\mathcal{T}_{S'C}(H_{S'C}) = H_{S'C}$ and $\mathcal{T}_{S'C}(\hat{1}_{S'C}) = \hat{1}_{S'C}$ implies $\mathcal{T}_{S'C}(e^{\alpha H_{S'C}}) = e^{\alpha H_{S'C}}$ for all $\alpha \in \mathbb{C}$. Moreover, due to the orthogonal supports of $H^i_{S'} \otimes P^i_C$, $H^f_{S'} \otimes P^f_C$, and $H^\perp$ we can conclude that

$$\begin{aligned} e^{-\beta H^i_{S'}} \otimes |c_{i\pm}\rangle\langle c_{i\pm}| =& e^{-\beta H_{S'C}}[\hat{1}_{S'} \otimes |c_{i\pm}\rangle\langle c_{i\pm}|] \\ =& [\hat{1}_{S'} \otimes |c_{i\pm}\rangle\langle c_{i\pm}|]e^{-\beta H_{S'C}}. \end{aligned} \tag{C6}$$

We examplify the rest of the derivation with the transformation of $G_\beta(H^i_{S'}) \otimes |c_{i+}\rangle\langle c_{i+}|$.

$$\begin{aligned} &\mathcal{T}_{S'C}\big(G(H^i_{S'}) \otimes |c_{i+}\rangle\langle c_{i+}|\big) \\ &\text{[By Eq. (C6)]} \\ =& \frac{1}{Z(H^i_{S'})}\mathcal{T}_{S'C}(e^{-\beta H_{S'C}}[\hat{1}_{S'} \otimes |c_{i+}\rangle\langle c_{i+}|]) \\ &\text{[By property (B3a) and } \mathcal{T}_{S'C}(e^{\alpha H_{S'C}}) = e^{\alpha H_{S'C}}] \\ =& \frac{1}{Z(H^i_{S'})}\mathcal{T}_{S'C}(\hat{1}_{S'} \otimes |c_{i+}\rangle\langle c_{i+}|)e^{-\beta H_{S'C}} \\ =& \frac{1}{Z(H^i_{S'})}[\hat{1}_{S'} \otimes |c_{i-}\rangle\langle c_{i-}|]e^{-\beta H_{S'C}} \\ &\text{[By Eq. (C6)]} \\ =& G(H^i_{S'}) \otimes |c_{i-}\rangle\langle c_{i-}|, \end{aligned}$$

The other identities can be derived in an analogous manner. □

2. The induced channels

Given the initial state $|c_{i+}\rangle$ and final state $|c_{f+}\rangle$ we define the channels

$$\begin{aligned} \mathcal{F}_+(\sigma) &:= \mathrm{Tr}_{S'C}(V[G_\beta(H^i_{S'}) \otimes |c_{i+}\rangle\langle c_{i+}| \otimes \sigma]V^\dagger), \\ \mathcal{R}_+(\sigma) &:= \mathrm{Tr}_{S'C}(V^\dagger[G_\beta(H^f_{S'}) \otimes |c_{f+}\rangle\langle c_{f+}| \otimes \sigma]V). \end{aligned} \tag{C7}$$

For the initial state $|c_{f-}\rangle$ and the final state $|c_{i-}\rangle$ we similarly define the channels

$$\begin{aligned} \mathcal{F}_-(\sigma) &:= \mathrm{Tr}_{S'C}(V[G_\beta(H^f_{S'}) \otimes |c_{f-}\rangle\langle c_{f-}| \otimes \sigma]V^\dagger), \\ \mathcal{R}_-(\sigma) &:= \mathrm{Tr}_{S'C}(V^\dagger[G_\beta(H^i_{S'}) \otimes |c_{i-}\rangle\langle c_{i-}| \otimes \sigma]V). \end{aligned} \tag{C8}$$

3. Deriving a quantum Crooks relation

Lemma 11. *Given Assumptions 2, then the channels $\mathcal{R}_+$, $\mathcal{F}_-$, as defined in Eqs. (C7) and (C8) are related as*

$$\mathcal{T}_E\mathcal{R}_+ = \mathcal{F}_-\mathcal{T}_E \tag{C9}$$

and thus

$$\mathcal{R}^*_+ = \mathcal{F}^\ominus_-. \tag{C10}$$

Hence, under the assumption of time-reversal symmetry we can in effect simulate the reversed time evolution (i.e., the replacement of V with $V^\dagger$) via the 'forward' evolution V. By applying the property $\mathcal{T}_E^2 = I$ to (C9) one can also show $\mathcal{R}_+\mathcal{T}_E = \mathcal{T}_E\mathcal{F}_-$ and $\mathcal{R}_+ = \mathcal{T}_E\mathcal{F}_-\mathcal{T}_E$.

Proof of Lemma 11. We use the definition of $\mathcal{R}_+$ in Eq. (C7) and the general relation $\mathrm{Tr}_2\big([\mathcal{T}_1 \otimes \mathcal{T}_2](\rho)\big) = \mathcal{T}_1\big(\mathrm{Tr}_2(\rho)\big)$ to obtain

$$\begin{aligned}
&\mathcal{T}_E \circ \mathcal{R}_+(\sigma)\\
=&\mathrm{Tr}_{S'C}\Big(\mathcal{T}\big(V^\dagger[G(H^f_{S'}) \otimes |c_{f+}\rangle\langle c_{f+}| \otimes \sigma]V\big)\Big)\\
&\text{[By (B3a), (B3b), and } \mathcal{T}(V) = V]\\
=&\mathrm{Tr}_{S'C}\Big(V\mathcal{T}\big(G(H^f_{S'}) \otimes |c_{f+}\rangle\langle c_{f+}| \otimes \sigma\big)V^\dagger\Big) \qquad \text{(C11)}\\
&\text{[By Lemma 10]}\\
=&\mathrm{Tr}_{S'C}\big(V[G(H^f_{S'}) \otimes |c_{f-}\rangle\langle c_{f-}| \otimes \mathcal{T}_E(\sigma)]V^\dagger\big)\\
=&\mathcal{F}_- \circ \mathcal{T}_E(\sigma).
\end{aligned}$$

By multiplying (C9) from the left with $\mathcal{T}$ and using the relation $\mathcal{T}^2 = I$ and the the alternative definition of $\ominus$ in (B5) we obtain (C10). □

Proposition 3 (Quantum Crooks relation)**.** *With Assumptions 2, the channels $\mathcal{F}_\pm$ as defined in Eqs. (C7) and (C8) satisfy*

$$Z(H^i_{S'})\mathcal{F}_+ = Z(H^f_{S'})\mathcal{J}_{\beta H_E}\mathcal{F}_-^{\ominus}\mathcal{J}^{-1}_{\beta H_E}. \qquad \text{(C12)}$$

With the separation of S' into system S and the heat bath B as in equation (A2) we thus get

$$Z(H^i_S)\mathcal{F}_+ = Z(H^f_S)\mathcal{J}_{\beta H_E}\mathcal{F}_-^{\ominus}\mathcal{J}^{-1}_{\beta H_E}. \qquad \text{(C13)}$$

Proof. The triple V, $|c_i\rangle := |c_{i+}\rangle$, $|c_f\rangle := |c_{f+}\rangle$ from Assumptions 2 satisfies Assumptions 1. It follows that we can apply Proposition 1 on the pair of channels $\mathcal{F}_+$ and $\mathcal{R}_+$ and thus obtain $Z(H^i_{S'})\mathcal{F}_+ = Z(H^f_{S'})\mathcal{J}_{\beta H_E}\mathcal{R}_+^*\mathcal{J}^{-1}_{\beta H_E}$. Next, we use Eq. (C10) to obtain Eq. (C12).

With the additional assumption in equation (A2) we get $Z(H^i_{S'}) = Z(H^i_{S'})Z(H_B)$ and $Z(H^f_{S'}) = Z(H^f_{S'})Z(H_B)$. From this follows that (C12) yields (C13). □

4. Unbounded H_E

The requirement of perfect control, i.e., that V satisfies (C3) puts rather stringent conditions on the properties of H_E. To see this, let us assume that H_E has a pure point spectrum corresponding to the complete orthonormal basis $\{|n\rangle\}_n$ with corresponding energy eigenvalues E_n. Let us furthermore assume that $\mathcal{H}_{S'}$ is finite-dimensional. Here we shall see that for generic choices of initial and final Hamiltonians $H^i_{S'}$ and $H^f_{S'}$, the perfect control implies that the spectrum of H_E must be unbounded from both above and below.

Due to the assumption of energy conservation, the energy reservoir has to compensate for any change in energy in the transition from the initial to the final state. Let h^i_n be the eigenvalues of $H^i_{S'}$ and similarly h^f_m the eigenvalues of $H^f_{S'}$. Suppose that $h^f_m \neq h^i_n$ for all m, n. This means that every possible transition either must cost or yield energy, which has to be drawn from or deposited in the reservoir E. Imagine now that S' initially is in an eigenstate $|h^i_n\rangle$. Suppose that at the end of the process there is a non-zero probability for finding S' in the state $|h^f_m\rangle$ with $h^f_m > h^i_m$. For this to happen, the reservoir has to donate the energy $q := h^i_m - h^f_n$. Suppose that the spectrum of H_E would be bounded from below, i.e., $E_{\text{lower}} = \inf_n E_n > -\infty$. This means that there exists some state $|k\rangle$ of the reservoir such that all transitions downwards in energy (if any available) would be smaller than q. In other words, if the energy reservoir would start in state $|k\rangle$, then it cannot donate the energy q, and the transition cannot occur. For a reservoir with a spectrum bounded from below, the only way to avoid this would be if all transitions in S' always would go downwards in energy. Generic choices of $H^i_{S'}$ to $H^f_{S'}$ would involve both increases and decreases in energy, and thus the spectrum of H_E must be unbounded from both above and below. The key point behind the unboundedness is the demand that the control system always should succeed in its task irrespective of the state of the system and the energy reservoir. It would be reasonable with a control system fails in some cases, e.g., if the energy in the reservoir is too low (i.e. too close to the ground state). In section H we introduce conditional fluctuation relations that allows for failing control systems. (For an explicit example, see section H 7 a.)

Appendix D: Diagonal and off-diagonal Crooks relations

1. Decoupling of diagonals

We here demonstrate the useful fact that the dynamics under the induced channels $\mathcal{F}_\pm$ and $\mathcal{R}_\pm$ decouples along different diagonals or modes of coherence [34]. We first show that the channels $\mathcal{F}_\pm$ and $\mathcal{R}_\pm$ commute with the commutator with respect to H_E.

Lemma 12. *With Assumptions 2, the channels $\mathcal{F}_\pm$ and $\mathcal{R}_\pm$ as defined in Eqs. (C7) and (C8), satisfy*

$$\begin{aligned}
[H_E, \mathcal{F}_\pm(\sigma)] =&\mathcal{F}_\pm([H_E, \sigma]),\\
[H_E, \mathcal{R}_\pm(\sigma)] =&\mathcal{R}_\pm([H_E, \sigma]).
\end{aligned} \qquad \text{(D1)}$$

Proof. Here we only show the relation $[H_E, \mathcal{F}_+(\sigma)] = \mathcal{F}_+([H_E, \sigma])$. By the definition of the Hamiltonian H in Assumptions 2 it follows that $\hat{1}_{S'} \otimes |c_{f+}\rangle\langle c_{f+}| \otimes H_E = (H - H^f_{S'} \otimes |c_{f+}\rangle\langle c_{f+}| \otimes \hat{1}_E)[\hat{1}_{S'} \otimes |c_{f+}\rangle\langle c_{f+}| \otimes \hat{1}_E]$. By combining this observation with the perfect control (C3)

one can show that

$$\begin{aligned}&[H_E, \mathcal{F}_+(\sigma)]\\ =&\mathrm{Tr}_{S'C}\Big(\Big[H, V[G_\beta(H^i_{S'})\otimes|c_{i+}\rangle\langle c_{i+}|\otimes\sigma]V^\dagger\Big]\Big)\\ &-\mathrm{Tr}_{S'C}\Big(\Big[H^f_{S'}\otimes|c_{f+}\rangle\langle c_{f+}|\otimes\hat{1}_E,\\ &\qquad V[G_\beta(H^i_{S'})\otimes|c_{i+}\rangle\langle c_{i+}|\otimes\sigma]V^\dagger\Big]\Big),\end{aligned} \tag{D2}$$

where the last term becomes zero due to the cyclic property of the partial trace $\mathrm{Tr}_{S'C}$ with respect to $H^f_{S'}\otimes|c_{f+}\rangle\langle c_{f+}|\otimes\hat{1}_E$. By the definition of the global Hamiltonian H in Assumptions 2

$$\begin{aligned}&[H, G_\beta(H^i_{S'})\otimes|c_{i+}\rangle\langle c_{i+}|\otimes\sigma]\\ &=[H^i_{S'}, G_\beta(H^i_{S'})]\otimes|c_{i+}\rangle\langle c_{i+}|\otimes\sigma\\ &\quad+G_\beta(H^i_{S'})\otimes|c_{i+}\rangle\langle c_{i+}|\otimes[H_E,\sigma]\\ &=G_\beta(H^i_{S'})\otimes|c_{i+}\rangle\langle c_{i+}|\otimes[H_E,\sigma]\end{aligned}$$

By combining this with $[H,V]=0$ in (D2) the lemma follows. □

Corollary 1. *Suppose that H_E has a pure point spectrum, i.e., there exists a complete orthonormal basis $\{|n\rangle\}_n$ of eigenvectors of H_E with corresponding eigenvalues E_n. Then*

$$\begin{aligned}&\langle m|\mathcal{F}_\pm(|n\rangle\langle n'|)|m'\rangle=0\\ &\quad\text{if}\quad E_m-E_n\neq E_{m'}-E_{n'}.\end{aligned} \tag{D3}$$

The analogous statement holds for $\mathcal{R}_\pm$.

Proof. By Lemma 12

$$\begin{aligned}&(E_m-E_{m'})\langle m|\mathcal{F}_\pm(|n\rangle\langle n'|)|m'\rangle\\ &=\langle m|[H,\mathcal{F}_\pm(|n\rangle\langle n'|)]|m'\rangle\\ &=\langle m|\mathcal{F}_\pm([H,|n\rangle\langle n'|])|m'\rangle\\ &=(E_n-E_{n'})\langle m|\mathcal{F}_\pm(|n\rangle\langle n'|)|m'\rangle.\end{aligned}$$

Thus $(E_m-E_{m'}-E_n+E_{n'})\langle m|\mathcal{F}_\pm(|n\rangle\langle n'|)|m'\rangle=0$. □

If H_E is non-degenerate (i.e. $E_n=E_{n'}$ if and only if $n=n'$) then it follows by Corollary 1 that $\langle m|\mathcal{F}_\pm(Q)|m\rangle=\sum_n\langle m|\mathcal{F}_\pm\big(|n\rangle\langle n|Q|n\rangle\langle n|\big)|m\rangle$.

Hence, $\mathcal{F}_\pm$ cannot 'create' off-diagonal elements with respect to the energy eigenbasis. Moreover, if we are only interested in the diagonal elements of the output, we only need to consider the diagonal elements of the input. Another way to put it is that the statistics of an energy measurement on the output is unaffected by an additional energy measurement on the input.

2. Diagonal Crooks relations

Let us assume that H_E has a pure non-degenerate point spectrum with eigenenergies E_n corresponding to the complete orthonormal eigenbasis $\{|n\rangle\}_n$. We furthermore assume that $\mathcal{T}_E(|n\rangle\langle n|)=|n\rangle\langle n|$.

Imagine now that we represent the density operator of the energy reservoir as a matrix with respect to the basis $\{|n\rangle\}_n$. Since $\mathcal{F}_\pm$ are channels it follows that the numbers

$$p_\pm(m|n):=\langle m|\mathcal{F}_\pm(|n\rangle\langle n|)|m\rangle, \tag{D4}$$

can be interpreted as conditional probability distributions.

Proposition 4. *With Assumptions 2, a non-degenerate H_E with pure point spectrum, and a time-reversal $\mathcal{T}_E$ acting as the transpose with respect to the eigenbasis $\{|n\rangle\}_n$, the conditional distributions $p_+(m|n)$ and $p_-(n|m)$ defined in (D4) satisfy*

$$Z(H^i_S)p_+(m|n)=Z(H^f_S)e^{\beta(E_n-E_m)}p_-(n|m). \tag{D5}$$

The existence of the diagonal Crooks relation (H12) does strictly speaking not rely on the decoupling of the diagonals (cf. section H 5 where we do obtain a diagonal Crooks relation in a setting where there is no decoupling along the diagonals).

Proof. If we apply both sides of equation (C13) on the operator $|n\rangle\langle n|$, and operate on both sides of the resulting equality with $\langle m|\cdot|m\rangle$ we obtain

$$\begin{aligned}&Z(H^i_S)\langle m|\mathcal{F}_+(|n\rangle\langle n|)|m\rangle\\ &=Z(H^f_S)e^{\beta(E_n-E_m)}\langle m|\mathcal{F}^{\ominus}_-(|n\rangle\langle n|)|m\rangle.\end{aligned}$$

With the invariance of $|n\rangle\langle n|$ under the time-reversal, we find $\langle m|\mathcal{F}^{\ominus}_-(|n\rangle\langle n|)|m\rangle=\langle n|\mathcal{F}_-(|m\rangle\langle m|)|n\rangle$. With the identifications in (D4) the proposition follows. □

In section G 1 we shall use the additional assumption of energy translation invariance on the energy reservoir to show how (H12) leads to the standard classical Crooks relation.

3. Off-diagonal Crooks relations

Like in section D 2 we here assume a discrete non-degenerate spectrum of H_E, with corresponding orthonormal eigenbasis $\{|n\rangle\}_n$ and energy eigenvalues E_n. We also assume that $\mathcal{T}_E$ is the transpose with respect to this basis, and thus $\mathcal{T}_E(|n\rangle\langle n'|)=|n'\rangle\langle n|$.

As discussed in section D 1, the channel $\mathcal{F}_+$ can only induce transitions between $|n\rangle\langle n'|$ and $|m\rangle\langle m'|$ if $E_n-E_{n'}=E_m-E_{m'}$. For each δ we can thus define a corresponding set of operators $\{|n\rangle\langle n'|\}_{n,n':E_n-E_{n'}=\delta}$. (This set would be empty for many values of δ.) For each such δ, we will construct a Crooks relation, analogous to what we did for the diagonal case.

As the generalization of $p_+(m|n)$ and $p_-(n|m)$ we define

$$q^\delta_\pm(m|n):=\langle m|\mathcal{F}_\pm(|n\rangle\langle n'|)|m'\rangle, \tag{D6}$$

where $q^0_\pm = p_\pm$. The reason for why it is enough to write '$q^\delta_\pm(m|n)$' rather than '$q^\delta_\pm(mm'|nn')$' is that m' and n' are uniquely determined by δ, m, and n, due to the assumption that H_E is non-degenerate.

The set of numbers $q^\delta_\pm(m|n)$ represent the channels $\mathcal{F}_\pm$ in the sense that

$$\mathcal{F}_\pm(\rho) = \sum_\delta \sum_{m,n} \sum_{\substack{n'm':E_n-E_{n'}=\delta \\ E_m-E_{m'}=\delta}} q^\delta_\pm(m|n)|m\rangle\langle n|\rho|n'\rangle\langle m'|. \tag{D7}$$

Proposition 5. *With Assumptions 2, a non-degenerate H_E with pure point spectrum, and a time-reversal $\mathcal{T}_E$ acting as the transpose with respect to the eigenbasis $\{|n\rangle\}_n$, the $q^\delta_\pm(m|n)$ defined in (D6) satisfy*

$$Z(H^i_S)q^\delta_+(m|n) = Z(H^f_S)e^{\beta(E_n-E_m)}q^\delta_-(n|m). \tag{D8}$$

Proof. Let n, n' and m, m' be such that $E_n - E_{n'} = E_m - E_{m'} = \delta$. If we apply both sides of equation (C13) on the operator $|n\rangle\langle n'|$, and operate on both sides of the resulting equality with $\langle m| \cdot |m'\rangle$ we obtain

$$\begin{aligned} &Z(H^i_S)\langle m|\mathcal{F}_+(|n\rangle\langle n'|)|m'\rangle \\ &= Z(H^f_S)e^{\beta(E_n-E_m)}\langle m|\mathcal{F}^\ominus_-(|n'\rangle\langle n|)|m'\rangle, \end{aligned}$$

where we have made use of $E_{n'} = E_n - \delta$ and $E_{m'} = E_m - \delta$. With the assumption that $\mathcal{T}_E$ is the transpose with respect to $\{|n\rangle\}_n$, together with the identifications in equation (D6), we obtain the proposition. □

As mentioned in the main text we need to use off-diagonal initial states as well as off-diagonal measurement operators in order to determine the numbers $q^\delta_\pm(m|n)$ in a 'prepare and measure'- experiment. There are many possible arrangements, but let us here construct a setup that determines these numbers via interference. Let n, n' and m, m' be such that $\delta = E_n - E_{n'} = E_m - E_{m'}$. Define the POVM element $A := (|m\rangle + |m'\rangle)(\langle m| + \langle m'|)/2$ and the family of initial states $|\psi_\theta\rangle := (|n\rangle + e^{i\theta}|n'\rangle)/\sqrt{2}$. Then the probability of measuring A on the evolved state is

$$\begin{aligned} &\mathrm{Tr}\big(A\mathcal{F}_\pm(|\psi_\theta\rangle\langle\psi_\theta|)\big) \\ =&\frac{1}{4}\big(p_\pm(m|n) + p_\pm(m'|n) + p_\pm(m|n') + p_\pm(m'|n')\big) \\ &+ |q^\delta(m|n)|\cos\Big(\arg\big(q^\delta_\pm(m|n)\big) - \theta\Big), \end{aligned}$$

where we have made use of the decoupling. Hence, both the magnitude and phase of $q^\delta(m|n)$ can be determined via the amplitude and phase-shift in the interference pattern with respect to the phase shift θ.

Appendix E: Jarzynski equalities

Jarzynski's equality [80] can be formulated as $\langle e^{-\beta W}\rangle = Z(H^f)/Z(H^i)$. This is often written in the more elegant form $\langle e^{-\beta(W-\Delta F)}\rangle = 1$, where $\Delta F = F(H^f) - F(H^i)$, with $F(H) = -kT\ln Z(H)$, is the free-energy difference between the initial and final state. Here we obtain the following family of quantum Jarzynski equalities

Proposition 6. *With Assumptions 2, the channels $\mathcal{F}_+$ and $\mathcal{R}_+$ as defined in Eq. (C7) satisfy*

$$\begin{aligned} &\mathrm{Tr}\Big[e^{\beta H_E}\mathcal{F}_+\Big(e^{(-\beta+r+z)H_E/2}\rho e^{(-\beta+r-z)H_E/2}\Big)\Big] \\ &= \frac{Z(H^f_{S'})}{Z(H^i_{S'})}\mathrm{Tr}[e^{rH_E/2}\mathcal{R}_+(\hat{1})e^{rH_E/2}\rho], \end{aligned} \tag{E1}$$

for $r \in \mathbb{R}$ and $z \in \mathbb{C}$.

Hence, if $\mathcal{R}_+(\hat{1}_E) = \hat{1}_E$, then

$$\begin{aligned} &\mathrm{Tr}\Big[e^{\beta H_E}\mathcal{F}_+\Big(e^{(-\beta+r+z)H_E/2}\rho e^{(-\beta+r-z)H_E/2}\Big)\Big] \\ &= \frac{Z(H^f_{S'})}{Z(H^i_{S'})}\mathrm{Tr}(e^{rH_E}\rho), \end{aligned} \tag{E2}$$

The condition $\mathcal{R}_+(\hat{1}_E) = \hat{1}_E$ is equivalent to $\mathcal{F}_-(\hat{1}_E) = \hat{1}_E$, with $\mathcal{F}_-$ as defined in Eq. (C8).

The relations (14) and (15) in the main text are obtained as special cases of (E2).

Strictly speaking, the traces in the above expressions are not necessarily well defined and finite for all operators ρ. However, we here proceed under the assumption that ρ are chosen such that the traces are well defined.

Proof. By Assumptions 2 it follows that Proposition 3 is applicable, and thus the channels $\mathcal{F}_\pm$ as defined in Eqs. (C7) and (C8) satisfy equation (C12). By applying (C12) to the operator $e^{(-\beta+r+z)H_E/2}\rho e^{(-\beta+r-z)H_E/2}$, multiplying both sides of the resulting equality with $e^{\beta H_E}$, and take the trace, and divide by $Z(H^i_{S'})$, one obtains

$$\begin{aligned} &\mathrm{Tr}\Big[e^{\beta H_E}\mathcal{F}_+\big(e^{(-\beta+r+z)H_E/2}\rho e^{(-\beta+r-z)H_E/2}\big)\Big] \\ &= \frac{Z(H^f_{S'})}{Z(H^i_{S'})}\mathrm{Tr}\Big[\mathcal{F}^\ominus_-\Big(e^{(r+z)H_E/2}\rho e^{(r-z)H_E/2}\Big)\Big] \\ &\text{[By Eq. (C10) and the definition of '}*\text{']} \\ &= \frac{Z(H^f_{S'})}{Z(H^i_{S'})}\mathrm{Tr}\Big[\mathcal{R}_+(\hat{1})e^{(r+z)H_E/2}\rho e^{(r-z)H_E/2}\Big] \end{aligned}$$

With the definition of the commutator $\mathcal{C}_{H_E}(\sigma) := [H_E, \sigma]$, we can write $e^{-z\mathcal{C}_{H_E}/2}\big(\mathcal{R}_+(\hat{1})\big) = e^{-zH_E/2}\mathcal{R}_+(\hat{1})e^{zH_E/2}$. Combined with the fact from Lemma 12, that $\mathcal{C}_{H_E}$ and $\mathcal{R}_+$ commute, we thus get $e^{-zH_E/2}\mathcal{R}_+(\hat{1})e^{zH_E/2} = \mathcal{R}_+(\hat{1})$. This proves (E1).

That $\mathcal{F}_-(\hat{1}) = \hat{1}$ if and only in $\mathcal{R}_+(\hat{1}) = \hat{1}$, follows from Lemma 11 together with the properties $\mathcal{T}_E(\hat{1}) = \hat{1}$ (Lemma 3) and $\mathcal{T}_E^2 = I$. □

One may get the impression that the members in the family of equalities in Proposition 6 are independent. However, at least in the finite-dimensional case one can transform them into each other, and in this sense they should maybe rather be regarded as the same equality in different guises. To see this, start with assuming that (E1) is true for all operators ρ on a finite-dimensional Hilbert space. We wish to show that this implies that (E1) is also true for r and z substituted with arbitrary r' and z'. Let $r = r' + \Delta r$ and $z = z' + \Delta z$ in (E1) and define $\rho' := e^{\Delta r H_E/2}\rho e^{\Delta r H_E/2}$. This yields the equality

$$\begin{aligned}&\mathrm{Tr}\Big[e^{\beta H_E}\mathcal{F}_+\Big(e^{(-\beta+r'+z'+\Delta z)H_E/2}\rho' e^{(-\beta+r'-z'-\Delta z)H_E/2}\Big)\Big]\\ &= \frac{Z(H^f_{S'})}{Z(H^i_{S'})}\mathrm{Tr}[e^{r'H_E/2}\mathcal{R}_+(\hat{1})e^{r'H_E/2}\rho'],\end{aligned}$$

We can now use the fact (Lemma 12) that the commutator $\mathcal{C}_{H_E}(\sigma) := [H_E, \sigma]$ commutes with $\mathcal{F}_+$, to show that $\mathcal{F}_+\big(e^{(-\beta+r'+z'+\Delta z)H_E/2}\rho' e^{(-\beta+r'-z'-\Delta z)H_E/2}\big) = e^{\Delta z H_E/2}\mathcal{F}_+\big(e^{(-\beta+r'+z')H_E/2}\rho' e^{(-\beta+r'-z')H_E/2}\big)e^{-\Delta z H_E/2}$. From this is follows that (E1) remains valid with r, z, ρ substituted by r', z', ρ'. In the finite-dimensional case it is also clear that the mapping $\rho \mapsto \rho' = e^{\Delta r H_E/2}\rho e^{\Delta r H_E/2}$ for Hermitian H_E and real Δr is a bijection on the space of linear operators. Hence, (E1) with r', z' holds for all operators ρ.

Appendix F: Bound on the work cost

For processes that start in equilibrium one would expect that the work cost should be bounded from below by the equilibrium free energy difference of the final and initial Hamiltonian of the system $\langle W\rangle \geq F(H^f) - F(H^i)$. Here we derive a similar expression in our setting, under the assumption $\mathcal{R}_+$ is unital, and that H_E has a pure non-degenerate point spectrum, i.e., that there exists a complete orthonormal basis $\{|n\rangle\}_n$ of eigenvectors to H_E corresponding to distinct eigenvalues E_n. (The latter assumptions may not necessarily be essential.)

1. Bound on the average energy loss in the reservoir

Proposition 7. *Let H_E have a pure non-degenerate point spectrum. Assume that the initial state is $G(H^i_{S'})\otimes\sigma$, with σ a density operator on $\mathcal{H}_E$. Then*

$$\begin{aligned}&\mathrm{Tr}(H_E\sigma) - \mathrm{Tr}\big(H_E\mathcal{F}_+(\sigma)\big)\\ &\geq F(H^f_{S'}) - F(H^i_{S'}) - \frac{1}{\beta}\ln\mathrm{Tr}(\sigma\mathcal{R}_+(\hat{1}_E)).\end{aligned} \tag{F1}$$

Hence, if $\mathcal{R}_+(\hat{1}) = \hat{1}$, then

$$\mathrm{Tr}(H_E\sigma) - \mathrm{Tr}\big(H_E\mathcal{F}_+(\sigma)\big) \geq F(H^f_{S'}) - F(H^i_{S'}). \tag{F2}$$

Hence, if we identify the loss of average energy in the energy reservoir, $\mathrm{Tr}(H_E\sigma) - \mathrm{Tr}\big(H_E\mathcal{F}_+(\sigma)\big)$, with $\langle W\rangle$, equation (F2) thus gives the standard bound.

Proof. Let E_n and $|n\rangle$ be eigenvalues and corresponding orthonormal eigenvectors to H_E such that $\{|n\rangle\}_n$ is a complete orthonormal basis to $\mathcal{H}_E$. Define

$$p_{n,m} := \langle n|\sigma|n\rangle\langle m|\mathcal{F}_+(|n\rangle\langle n|)|m\rangle. \tag{F3}$$

By the fact that $\mathcal{F}_+$ is trace preserving it follows that $\{p_{n,m}\}_{n,m}$ is a probability distribution. Corollary 1 and the non-degeneracy of H_E yield

$$\begin{aligned}&\mathrm{Tr}[e^{\beta H_E}\mathcal{F}_+(e^{-\beta H_E/2}\sigma e^{-\beta H_E/2})]\\ &= \sum_m\sum_n \langle n|\sigma|n\rangle e^{\beta(E_m-E_n)}\langle m|\mathcal{F}_+(|n\rangle\langle n|)|m\rangle\\ &\geq \exp\Big[\beta\sum_m\sum_n (E_m - E_n)\langle n|\sigma|n\rangle\langle m|\mathcal{F}_+(|n\rangle\langle n|)|m\rangle\Big]\\ &\quad[\text{By Corollary 1, and non-degenerate } H_E.]\\ &= \exp\Big[\beta\mathrm{Tr}(H_E\mathcal{F}_+(\sigma)) - \beta\mathrm{Tr}(\mathcal{F}_+(H_E\sigma))\Big]\\ &\quad[\mathcal{F}_+ \text{ is trace preserving}]\\ &= \exp\Big[\beta\mathrm{Tr}(H_E\mathcal{F}_+(\sigma)) - \beta\mathrm{Tr}(H_E\sigma)\Big],\end{aligned}$$

where the inequality follows by the convexity of the exponential function. By combining this inequality with the quantum Jarzynski equality (E1) in Proposition 6 for $r = 0$ and $z = 0$, one obtains

$$\frac{Z(H^f_S)}{Z(H^i_S)}\mathrm{Tr}[\sigma\mathcal{R}_+(\hat{1})] \geq e^{\beta\mathrm{Tr}(H_E\mathcal{F}_+(\sigma)) - \beta\mathrm{Tr}(H_E\sigma)}.$$

Since the logarithm is monotonically increasing, and thus preserves the inequality, we thus obtain (F1), where we use $F(H) = -kT\ln Z(H)$. □

2. An example

One should keep in mind that the inequality (F1) does not *per se* imply that the standard work bound $\langle W\rangle \geq F(H^f) - F(H^i)$ necessarily is violated when $\mathcal{R}_+$ is not unital; it only allows for the possibility. However, here we construct an explicit example where one indeed gets a violation.

With the setting as in Assumptions 2 we define the average energy loss $D(V,\sigma)$ in the energy reservoir

$$\begin{aligned}D(V,\sigma) &:= \mathrm{Tr}(H_E\sigma)\\ &\quad - \mathrm{Tr}\big([\hat{1}_{S'}\otimes\hat{1}_C\otimes H_E]V\rho_i V^\dagger\big),\\ \rho_i &:= G(H^i_{S'})\otimes|c_{i+}\rangle\langle c_{i+}|\otimes\sigma.\end{aligned} \tag{F4}$$

By energy conservation, the assumption of prefect control, and the general relation $H = F(H) - kT\ln G(H)$,

one can show that

$$D(V,\sigma) = F(H^f_{S'}) - F(H^i_{S'}) - \frac{1}{\beta}S\big(G(H^i_{S'})\big) - \frac{1}{\beta}\mathrm{Tr}\big([\ln G(H^f_{S'})\otimes \hat{1}_C\otimes \hat{1}_E]V\rho_i V^\dagger\big). \tag{F5}$$

The strategy will be to construct a model with a Hamiltonian and a particular class of energy conserving operators V that is simple enough that we can determine the corresponding minimum of $D(V,\sigma)$.

Let us now additionally assume that $H_E = s\sum_{j\in\mathbb{Z}} j|j\rangle\langle j|$, for an orthonormal basis $\{|j\rangle\}_{j\in\mathbb{Z}}$. (We use the very same Hamiltonian in section G.) We moreover assume that $\mathcal{H}_{S'}$ is finite-dimensional and that the eigenvalues of $H^i_{S'}$ and $H^f_{S'}$ are integer multiples of s, i.e., they have the eigenvalues $\{sz^i_n\}_{n=1}^N$, $\{sz^f_n\}_{n=1}^N$, for some $z^i_n, z^f_n \in \mathbb{Z}$, with corresponding eigenvectors $|\chi^i_n\rangle$, $|\chi^f_n\rangle$ (again the same as in section G). To make the derivations simpler, we also assume that these eigenvalues are non-degenerate. We define the following class of unitary operators on $\mathcal{H}_{S'CE}$

$$\begin{aligned} V :=& \sum_{j\in\mathbb{Z}}\sum_{n,m=1}^N U^{(j):f+,i+}_{m,n}|f^+_{j,m}\rangle\langle i^+_{j,n}| \\ &+ \sum_{j\in\mathbb{Z}}\sum_{n,m=1}^N U^{(j):i-,f-}_{n,m}|i^-_{j,n}\rangle\langle f^-_{j,m}| \\ &+ \sum_{j\in\mathbb{Z}}\sum_{n,m=1}^N U^{(j):i+,f+}_{n,m}|i^+_{j,n}\rangle\langle f^+_{j,m}| \\ &+ \sum_{j\in\mathbb{Z}}\sum_{n,m=1}^N U^{(j):f-,i-}_{m,n}|f^-_{j,m}\rangle\langle i^-_{j,n}|, \end{aligned} \tag{F6}$$

where $|i^+_{j,n}\rangle := |\chi^i_n\rangle|c_{i+}\rangle|j - z^i_n\rangle$, $|i^-_{j,n}\rangle := |\chi^i_n\rangle|c_{i-}\rangle|j - z^i_n\rangle$, $|f^+_{j,m}\rangle := |\chi^f_m\rangle|c_{f+}\rangle|j - z^f_m\rangle$, $|f^-_{j,m}\rangle := |\chi^f_m\rangle|c_{f-}\rangle|j - z^f_m\rangle$, and where the matrices $[U^{(j):f+,i+}_{m,n}]_{m,n}$, $[U^{(j):i-,f-}_{n,m}]_{n,m}$, $[U^{(j):i+,f+}_{n,m}]_{n,m}$, and $[U^{(j):f-,i-}_{m,n}]_{m,n}$ are unitary for each fixed j. By construction, V is energy conserving. (The class of energy translation invariant unitaries that we consider in section G is obtained if we additionally assume that these matrices are independent of j.)

Next we define time-reversals on $S'C$ and E. First define $Y := |c_{i+}\rangle\langle c_{i-}| + |c_{i-}\rangle\langle c_{i+}| + |c_{f+}\rangle\langle c_{f-}| + |c_{f-}\rangle\langle c_{f+}|$. We use this in turn to define $\mathcal{T}_{S'C}(Q) := [\hat{1}_{S'}\otimes Y]Q^t[\hat{1}_{S'}\otimes Y^\dagger]$, where t denotes the transpose with respect to the orthonormal basis $\mathcal{B} := \{|\chi^i_n\rangle|c_{i+}\rangle\}_n \cup \{|\chi^i_n\rangle|c_{i-}\rangle\}_n \cup \{|\chi^f_n\rangle|c_{f+}\rangle\}_n \cup \{|\chi^f_n\rangle|c_{f-}\rangle\}_n$. If this would have been a finite-dimensional case, we could have used Proposition 2 to conclude that $\mathcal{T}_{S'C}$ is a time-reversal. However, it is straightforward to directly check the properties in Definition 1.

Let $\mathcal{T}_E$ be the transpose with respect to the basis $\{|j\rangle\}_{j\in\mathbb{Z}}$ of $\mathcal{H}_E$, and define $\mathcal{T} := \mathcal{T}_{S'C}\otimes\mathcal{T}_E$. With this definition one can confirm that $\mathcal{T}(V) = V$ if and only if $U^{(j):i-,f-}_{n,m} = U^{(j):f+,i+}_{m,n}$ and $U^{(j):f-,i-}_{m,n} = U^{(j):i+,f+}_{n,m}$. Next, note that

$$\begin{aligned} &\mathrm{Tr}\big([\ln G(H^f_{S'})\otimes\hat{1}_C\otimes\hat{1}_E]V\rho_i V^\dagger\big) \\ &= \sum_{j\in\mathbb{Z}}\sum_{n,m=1}^N |U^{(j):f+,i+}_{m,n}|^2 \ln G_m(H^f_{S'}) \\ &\qquad\times G_n(H^i_{S'})\langle j - z^i_n|\sigma|j - z^i_n\rangle \end{aligned} \tag{F7}$$

In order to minimize (F5) over the time-reversal symmetric operators V in our designated family (F6) it is sufficient to minimize over the collection of unitary matrices $U^{(j):f+,i+} := [U^{(j):f+,i+}_{m,n}]_{m,n}$. (Since $\mathcal{T}(V) = V$ if and only if $U^{(j):f+,i+}_{m,n} = U^{(j):i-,f-}_{n,m}$ and $U^{(j):i+,f+}_{n,m} = U^{(j):f-,i-}_{m,n}$, there are no further restrictions.) Next, insert (F6) into (F5) and minimize, which yields

$$\min_{\{V\}_j} D(V,\sigma) = F(H^f_{S'}) - F(H^i_{S'}) - \frac{1}{\beta}S\big(G(H^i_{S'})\big) - \frac{1}{\beta}\sum_j\sum_m \lambda^\downarrow_m\Big(r_j\big(G(H^i_{S'})\big)\Big)\ln\lambda^\downarrow_m\big(G(H^f_{S'})\big), \tag{F8}$$

where

$$r^{(j)}(\rho) := \sum_n \langle j - z^i_n|\sigma|j - z^i_n\rangle|\chi^i_n\rangle\langle\chi^i_n|\rho|\chi^i_n\rangle\langle\chi^i_n|, \tag{F9}$$

and where $\lambda^\downarrow_n(Q)$ denotes the n:th eigenvalue of Q, ordered non-increasingly. The minimum in (F8) can be obtained by noting that $[|U^{(j):f+,i+}_{m,n}|^2]_{m,n}$ is a doubly stochastic matrix for each j. Hence, according to Birkhoff's theorem [142] it can be regarded as a convex combination of permutation matrices. Since every permutation matrix results from a unitary matrix, we know that the maximum of (F7) is given by a permutation. (Alternatively, one can define $U^{(j)} := \sum_{m,n} U^{(j):f+,i+}_{m,n}|\chi^f_m\rangle\langle\chi^i_n|$ and observe that $\sum_{n,m=1}^N |U^{(j):f+,i+}_{m,n}|^2 \ln G_m(H^f_{S'})G_n(H^i_{S'})\langle j - z^i_n|\sigma|j - z^i_n\rangle = \mathrm{Tr}\big[U^{(j)\dagger}\ln G(H^f_{S'})U^{(j)}r_j\big(G(H^i_{S'})\big)\big]$, where $U^{(j)} := \sum_{m,n} U^{(j):f+,i+}_{m,n}|\chi^f_m\rangle\langle\chi^i_n|$. By the general relation $\max_U \mathrm{Tr}(U^\dagger QUR) = \sum_m \lambda^\downarrow_m(Q)\lambda^\downarrow_m(R)$, see e.g. Theorem 4.3.53 in [141], it follows that (F8) holds.)

Assume that the energy reservoir starts in the specific energy eigenstate $\sigma := |0\rangle\langle 0|$. By the definition of r_j in (F9) this assumption, together with the assumed non-degeneracy of $H^i_{S'}$ and thus of z^i_n, leads to

$$\begin{aligned} &\sum_j\sum_m \lambda^\downarrow_m\Big(r_j\big(G(H^i_{S'})\big)\Big)\ln\lambda^\downarrow_m\big(G(H^f_{S'})\big) \\ &= \sum_n\sum_m \lambda^\downarrow_m\big(G_n(H^i_{S'})|\chi^i_n\rangle\langle\chi^i_n|\big)\ln\lambda^\downarrow_m\big(G(H^f_{S'})\big) \\ &= \sum_n G_n(H^i_{S'})\ln\lambda^\downarrow_1\big(G(H^f_{S'})\big) \\ &= \ln\lambda^\downarrow_1\big(G(H^f_{S'})\big). \end{aligned}$$

By combining this observation with (F8) we get

$$\min_{\{U^{(j)}\}_j} D(V,|0\rangle\langle 0|) = F(H^f_{S'}) - F(H^i_{S'}) - \frac{1}{\beta}S(G(H^i_{S'})) - \frac{1}{\beta}\ln\lambda_1^{\downarrow}(G(H^f_{S'})). \tag{F10}$$

In other words, we get a violation of the standard bound whenever $S\big(G(H^i_{S'})\big) + \ln\lambda_1^{\downarrow}\big(G(H^f_{S'})\big) > 0$. For an explicit example where this is the case, let $H^i_{S'} = H^f_{S'} = s\sum_{k=0}^{K} k|\chi_k\rangle\langle\chi_k|$. In this particular case we find

$$\begin{aligned} &S\big(G(H^i_{S'})\big) + \ln\lambda_1^{\downarrow}\big(G(H^f_{S'})\big) \\ &= \frac{s\beta\sum_{k=0}^{K} ke^{-s\beta k}}{\sum_{k'=0}^{K} e^{-s\beta k'}} \\ &= \frac{s\beta e^{-s\beta}}{1-e^{-s\beta}} - \frac{s\beta(K+1)e^{-s\beta(K+1)}}{1-e^{-s\beta(K+1)}}. \end{aligned}$$

In the limit of large K this approaches $s\beta e^{-s\beta}/(1-e^{-s\beta}) > 0$ if $s\beta > 0$. Thus, for sufficiently large K it follows that $D(V,\sigma) < F(H^f_{S'}) - F(H^i_{S'})$. Hence, with the identification between $D(V,\sigma)$ and $\langle W\rangle$ we do get a violation of the standard bound.

Appendix G: Energy translation invariance

We have in this investigation allowed for the possibility that the processes depends non-trivially on the amount of energy in the energy reservoir. Here we consider a further restriction that implements the idea that the setup does not depend on this energy level. This model has previously been used in [31] to analyze coherence and work extraction. Here we only describe the most essential aspects of this model. For a more detailed description, see [31]. First of all, imagine the Hamiltonian H_E of the energy reservoir as a doubly infinite ladder of energy levels

$$H_E = s\sum_{j\in\mathbb{Z}} j|j\rangle\langle j| \tag{G1}$$

with energy spacing $s > 0$. (See also the continuum version in [44].) As one can see, this Hamiltonian has a bottomless spectrum (which echoes the discussions in Sec. C 4). Although this is not the most physically satisfying assumption, one can view it as an idealization of the assumption that the energy content of the 'battery' is much higher than the characteristic scale of energy costs in the experiment. We furthermore assume that the Hamiltonian $H_{\widetilde{S}}$ of system $\widetilde{S}$ (which includes all systems that are not E, i.e., in our case, system S, the heath bath B, and the control C) is such that all its eigenvalues (we assume a finite-dimensional Hilbert space $\mathcal{H}_{\widetilde{S}}$ with dimension N) are integer multiples of the energy spacing s. (Due to this assumption it becomes easy to construct non-trivial energy conserving unitary operations.) In other words, we assume that $H_{\widetilde{S}}$ has an eigenbasis $\{|\psi_n\rangle\}_{n=1}^N$ with corresponding eigenvalues sz_n where $z_n \in \mathbb{Z}$ for each n. Note that we allow $H_{\widetilde{S}}$ to be degenerate, in which case $\{|\psi_n\rangle\}_n$ is an eigenbasis of our choice.

In this section we not only demand that the global unitary operations are energy conserving, $[H,V] = 0$, but also that they are energy translation invariant. To define what we mean by this, we introduce the energy translation operator $\Delta = \sum_j |j+1\rangle\langle j|$ on the energy reservoir. We say that a unitary operator V on $\mathcal{H}_{\widetilde{S}}\otimes\mathcal{H}_E$ is energy translation invariant if $[\hat{1}_{\widetilde{S}}\otimes\Delta^a, V] = 0$ for all $a \in \mathbb{Z}$. It turns out [31] that that all energy conserving and energy translation invariant unitary operators in this model can be written in the following way

$$V(U) = \sum_{n,n'=1}^{N} |\psi_n\rangle\langle\psi_n|U|\psi_{n'}\rangle\langle\psi_{n'}| \otimes \Delta^{z_{n'}-z_n}, \tag{G2}$$

where U is an arbitrary unitary operator on $\mathcal{H}_{\widetilde{S}}$. If there are degeneracies in the Hamiltonian $H_{\widetilde{S}}$ then $V(U)$ is independent of the choice of energy eigenbasis $\{|\psi_n\rangle\}_n$. In particular, if $\{P_m\}_m$ is a collection of eigenprojectors of $H_{\widetilde{S}}$, then one can alternatively write

$$V(U) = \sum_{m,m'} P_m U P_{m'} \otimes \Delta^{z_{m'}-z_m}. \tag{G3}$$

A useful property of $V(U)$ is that it preserves products $V(U_2U_1) = V(U_2)V(U_1)$. (In contrast to the time-reversals $\mathcal{T}$, there is no swap of the ordering.)

We also need to incorporate time-reversals (an aspect not included in [31]).

Lemma 13. *Let $\mathcal{T}_E$ be defined as the transpose with respect to the orthonormal basis $\{|j\rangle\}_{j\in\mathbb{Z}}$. Let $\mathcal{H}_{\widetilde{S}}$ be finite-dimensional and let $\mathcal{T}_{\widetilde{S}}$ be such that $\mathcal{T}_{\widetilde{S}}(H_{\widetilde{S}}) = H_{\widetilde{S}}$. Then $\mathcal{T} := \mathcal{T}_{\widetilde{S}}\otimes\mathcal{T}_E$ satisfies*

$$\mathcal{T}\big(V(U)\big) = V\big(\mathcal{T}_{\widetilde{S}}(U)\big), \tag{G4}$$

for all operators U on $\mathcal{H}_{\widetilde{S}}$.

The proof is a direct application of the properties of time-reversals combined with Lemma 7 and (G3).

The following lemma shows that all actions on the energy reservoir in this model automatically satisfy the condition for unitality of the induced channels $\mathcal{R}_\pm$ and $\mathcal{F}_\pm$ that emerged in the considerations on Jarzynski relations and work bounds in sections E and F, respectively. (The unitality of this type of induced channels was previously observed in section II.C in the Supplementary Material of [31].)

Lemma 14. *For the channels $\mathcal{F}_\pm$ and $\mathcal{R}_\pm$ defined in equations (C7) and (C8), with $V := V(U)$ as in (G2), it is the case that $\mathcal{F}_\pm(\hat{1}_E) = \hat{1}_E$ and $\mathcal{R}_\pm(\hat{1}_E) = \hat{1}_E$.*

The proof is obtained by inserting the definition (G2) of $V(U)$ into the definitions of $\mathcal{F}_\pm$ and $\mathcal{R}_\pm$ in equations (C7) and (C8), and apply these to the identity operator.

Lemma 15. *With $\mathcal{F}_\pm$ and $\mathcal{R}_\pm$ as defined in equations (C7) and (C8), with $V := V(U)$ as in (G2), it is the case that*

$$\begin{aligned}\Delta^j \mathcal{F}_\pm(\sigma)\Delta^{\dagger k} &= \mathcal{F}_\pm\big(\Delta^j \sigma \Delta^{\dagger k}\big),\\ \Delta^j \mathcal{R}_\pm(\sigma)\Delta^{\dagger k} &= \mathcal{R}_\pm\big(\Delta^j \sigma \Delta^{\dagger k}\big).\end{aligned} \tag{G5}$$

Proof. Here we only show the equality for $\mathcal{F}_+$. The others are obtained analogously. The proof is based on the fact that $V(U)$ commutes with $\hat{1}_{\widetilde{S}} \otimes \Delta^j$. With the notation $\eta_{S'C} := G(H^i_{S'}) \otimes |c_{i+}\rangle\langle c_{i+}|$ we can write

$$\begin{aligned}\Delta^j \mathcal{F}_+(\sigma)\Delta^{\dagger k} &= \mathrm{Tr}_{S'C}\Big([\hat{1}_{\widetilde{S}} \otimes \Delta^j]V(U)\\ &\quad\times [\eta_{S'C}\otimes\sigma]V^\dagger(U)[\hat{1}_{\widetilde{S}} \otimes \Delta^{\dagger k}]\Big)\\ &= \mathrm{Tr}_{S'C}\big(V(U)[\eta_{S'C}\otimes \Delta^j\sigma\Delta^{\dagger k}]V^\dagger(U)\big)\\ &= \mathcal{F}_+(\Delta^j\sigma\Delta^{\dagger k}).\end{aligned}$$

□

We can regard $\langle m|\mathcal{F}_\pm(|n\rangle\langle n'|)|m'\rangle$ as the matrix elements in a matrix representation of the linear maps $\mathcal{F}_\pm$. It turns out that the translation invariance in Lemma 15 in conjunction with the decoupling between the coherence modes described in section D 1 reduces the number of independent parameters in this representation. More precisely,

$$\langle m|\mathcal{F}_\pm(|n\rangle\langle n'|)|m'\rangle = \begin{cases} p_\pm(m-n|0), & n-m = n'-m' \\ 0, & n-m \neq n'-m' \end{cases} \tag{G6}$$

where $p_\pm(m|n) = \langle m|\mathcal{F}_\pm(|n\rangle\langle n|)|m\rangle$ is the diagonal transition probabilities as defined in (D4). The mapping $\mathcal{F}_\pm$ is thus determined by the probabilities by which $|0\rangle\langle 0|$ is mapped to the other eigenstates of H_E. Equivalently, this can be expressed as

$$\mathcal{F}_\pm(\rho) = \sum_{n,n',k\in\mathbb{Z}} p_\pm(k|0)\langle n|\rho|n'\rangle |n+k\rangle\langle n'+k|. \tag{G7}$$

The expression in (G7) can be compared with the more general case in (D7). In section D 3 it was demonstrated that the off-diagonal modes of coherence satisfy Crooks relations that are structurally identical to the one along the diagonal. Equations (G6) and (G7) implies a stronger statement for the special case of the energy translation invariant model. Namely, that the dynamical map along each the off-diagonal modes is identical to the one along the main diagonal. (This does not imply that the elements of the density matrix along the different diagonals are the same.)

1. Regaining the standard Crooks and Jarzynski relations

The standard Crooks relation can be written

$$Z(H^i_S)P_+(w) = e^{\beta w} Z(H^f_S)P_-(-w), \tag{G8}$$

where $P_\pm(w) := P(W_\pm = w)$, W_+ and W_- being the work costs the the forward and reverse processes regarded as random variables. In the following we shall see how one can regain (G8) from the diagonal Crooks relation (H12) in section D 2 by additionally assuming energy translation invariance in the energy ladder model.

If we identify the loss of energy in the reservoir with the work done, then a change from energy E_j to $E_{j'}$ in the reservoir corresponds to the work $w = E_n - E_m$ (cf. [9]). The probability $P_\pm(w)$ is obtained by summing up the probabilities of all the transitions that have the work cost w. More precisely,

$$P_\pm(w) := \sum_{n,m:E_n-E_m=w} p_\pm(m|n)\langle n|\sigma|n\rangle, \tag{G9}$$

where $p_\pm(m|n)$ are the conditional probability distributions, defined in (D4), that describe the transitions among the diagonal elements.

The energy-ladder model yields $E_n = sn$. Hence, $n - m = w/s$. By (G6) it follows that

$$p_\pm(m|n) = p_\pm(m-n|0) = p_\pm(0|n-m). \tag{G10}$$

A direct consequence is that the probability distribution $P_\pm(w)$, defined in (G9), becomes independent of the initial σ,

$$P_\pm(w) = p_\pm(-w/s|0) = p_\pm(0|w/s). \tag{G11}$$

These observations can be used to regain the classical Crooks relation (G8).

$$\begin{aligned}Z(H^i_S)P_+(w) =& Z(H^i_S)p_+(0|w/s)\\ =& Z(H^f_S)e^{\beta w}p_-(w/s|0)\\ =& Z(H^f_S)e^{\beta w}P_-(-w).\end{aligned} \tag{G12}$$

where the second inequality is due to the diagonal Crooks relation in (H12). Here $\sum_w$ means that we sum over the set of possible energy changes, which for this particular model is $s\mathbb{Z}$.

Since we have re-derived the classical Crooks relation we more or less automatically also obtain the classical Jarzynski equality $\langle e^{-\beta W}\rangle = Z(H^f)/Z(H^i)$ [80]. This can be obtained via the 'standard' derivation $\sum_w e^{-\beta w}Z(H^i)P_+(w) = \sum_w Z(H^f)P_-(-w) = Z(H^f)$, where the second equality is due to the Crooks relation (G12). One can alternatively use the fact that $\mathcal{R}_+(\hat{1}_E) = \hat{1}_E$ (from Lemma 14) and use the quantum Jarzynski equality (E2) to derive the classical Jarzynski relation, by again making use of the energy translation invariance.

2. $F_- \mapsto F_-^\ominus$ as a generalization of $P_-(w) \mapsto P_-(-w)$

In the main text it is claimed that the mapping $F_- \mapsto F_-^\ominus$ can be regarded as a generalization of the map

$P_-(w) \mapsto P_-(-w)$. This generalization becomes evident for the the energy translation invariant model. Let $\mathcal{T}_E$ be the transpose with respect to $\{|n\rangle\}_n$. By the translation invariance in Lemma G5 it follows that

$$\langle m|\mathcal{F}_\pm^{\ominus}(|n\rangle\langle n'|)|m'\rangle = \langle -m|\mathcal{F}_\pm(|-n\rangle\langle -n'|)|-m'\rangle. \tag{G13}$$

Let us now identify the work loss $w = s(n-m)$ and the offset $\delta = s(n-n')$. By Lemma 1 it follows that the *non-zero* elements $\langle m|\mathcal{F}_\pm^{\ominus}(|n\rangle\langle n'|)|m'\rangle$ can be written $\langle 0|\mathcal{F}_\pm(|w/s\rangle\langle w/s+\delta/s|)|\delta/s\rangle$. By (G13) it follows that the non-zero coefficients satisfy

$$\begin{aligned}&\langle 0|\mathcal{F}_\pm^{\ominus}(|w/s\rangle\langle w/s+\delta/s|)|\delta/s\rangle\\ &= \langle 0|\mathcal{F}_\pm(|-w/s\rangle\langle -w/s-\delta/s|)|-\delta/s\rangle.\end{aligned}$$

Hence, both the work parameter w and the offset δ get reflected by the $\ominus$ operation. For the diagonal, $\delta = 0$, we can thus conclude that $P_\pm[\mathcal{F}_\pm^{\ominus}](w) = \langle 0|\mathcal{F}_\pm^{\ominus}(|w/s\rangle\langle w/s|)|0\rangle = \langle 0|\mathcal{F}_\pm(|-w/s\rangle\langle -w/s|)|0\rangle = P_\pm[\mathcal{F}_\pm](-w)$. Hence for the diagonal elements, the mapping $\ominus$ implements the transformation $P_-(w) \mapsto P_-(-w)$.

3. Examples

The main purpose of these examples is to show that there exist setups of Hamiltonians, states, and unitary operations that satisfy Assumptions 1 and Assumptions 2. We also take the opportunity to construct discretizations of paths of Hamiltonian within these models. The reason for why these demonstrations have been postponed until this section is that the energy-translation invariant systems have properties that make them convenient for constructing explicit examples. These examples are only sketched, and the details of the straightforward but sometimes rather long-winding confirmations are left to the reader.

a. *A minimal example without time-reversal*

Here we demonstrate a 'minimal' setup that satisfies the conditions in Assumptions 1. Let $H^i_{S'}$ and $H^f_{S'}$ be Hamiltonians on $\mathcal{H}_{S'}$ for which the eigenvalues are multiples of s, i.e., sz^i_n and sz^f_n for $z^i_n, z^f_n \in \mathbb{Z}$, and let $H_E := s\sum_j j|j\rangle\langle j|$ be the energy ladder. Let $|c_i\rangle, |c_f\rangle \in \mathcal{H}_C$ be two orthonormal states, and let

$$\begin{aligned}H_{S'C} :=& H^i_{S'} \otimes |c_i\rangle\langle c_i| + H^f_{S'} \otimes |c_f\rangle\langle c_f|,\\ H :=& H_{S'C} \otimes \hat{1}_E + \hat{1}_{SC} \otimes H_E,\end{aligned}$$

and $H^\perp = 0$. Let $U_{S'}$ be an arbitrary unitary operator on $\mathcal{H}_{S'}$ and define $U := U_{S'} \otimes |c_f\rangle\langle c_i| + \overline{U}_{S'} \otimes |c_i\rangle\langle c_f|$. (The unitary operator $\overline{U}_{S'}$ plays no direct role in the protocol, but is there to make U unitary.) As one can see, $U[\hat{1}_{S'} \otimes |c_i\rangle\langle c_i|] = U_{S'} \otimes |c_f\rangle\langle c_i| = [\hat{1}_{S'} \otimes |c_f\rangle\langle c_f|]U$. This U would in general not be energy conserving. However, the unitary operator $V(U)$ is by construction energy conserving on $S'CE$, i.e., $[H, V(U)] = 0$.

The operators $\hat{1}_{S'} \otimes |c_i\rangle\langle c_i|$ and $\hat{1}_{S'} \otimes |c_f\rangle\langle c_f|$ are block-diagonal with respect to the energy eigenspaces of $H_{S'C}$. Due to this one can confirm that

$$\begin{aligned}V(\hat{1}_{S'} \otimes |c_i\rangle\langle c_i|) =& \hat{1}_{S'} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E,\\ V(\hat{1}_{S'} \otimes |c_f\rangle\langle c_f|) =& \hat{1}_{S'} \otimes |c_f\rangle\langle c_f| \otimes \hat{1}_E.\end{aligned}$$

By combining this observation with the general property $V(A)V(B) = V(AB)$ and the perfect control of U, it follows that $V(U)[\hat{1}_{S'} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E] = [\hat{1}_{S'} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E]V(U)$. Hence, the fact that U satisfies perfect control implies that $V := V(U)$ also satisfies the condition for perfect control, and we can conclude that this setup satisfies all conditions of Assumptions 1.

b. *Discretized paths of Hamiltonians without time-reversal*

The intermediate Crooks relation only require us to consider the end-points of the dynamics. It may nevertheless be useful to see how one can construct a discretized model of a parametric family of Hamiltonians that satisfies Assumptions 1.

Given a family of Hamiltonians $H(x)$ for $x \in [0,1]$ with $H(0) = H^i_{S'}$ and $H(1) = H^f_{S'}$ we discretize the path into $L+1$ steps, such that we get a sequence of Hamiltonians $H_l := H(l/L)$ for $l = 0, \ldots, L$. Given the energy spacing s in the energy ladder, we find approximate Hamiltonians $\widetilde{H}_l$ that have the eigenvalues $sz^{(l)}_n$ for $z^{(l)}_n \in \mathbb{Z}$ with corresponding orthonormal eigenvectors $\{|\chi^{(l)}_n\rangle\}_n$. (For more details on the transition from H_l to $\widetilde{H}_l$, see Section VIII.C.1 in the Supplementary Material of [31].) We let $\{|c_l\rangle\}_{l=0}^L$ be a set of orthonormal elements spanning the Hilbert space $\mathcal{H}_C$ of the control system C, and define

$$\begin{aligned}H_{S'C} :=& \sum_{l=0}^{L} \widetilde{H}_l \otimes |c_l\rangle\langle c_l|,\\ H :=& H_{S'C} \otimes \hat{1}_E + \hat{1}_{SC} \otimes H_E,\end{aligned} \tag{G14}$$

for the energy ladder H_E. To compare with Assumptions 1 we have

$$\begin{aligned}&|c_i\rangle := |c_0\rangle, \quad |c_f\rangle := |c_L\rangle,\\ &H^i_{S'} := \widetilde{H}_0, \quad H^f_{S'} := \widetilde{H}_L,\\ &H^\perp := \sum_{l=1}^{L-1} \widetilde{H}_l \otimes |c_l\rangle\langle c_l|.\end{aligned} \tag{G15}$$

In the following we shall define a unitary operator U on $S'C$ that generates one single step along the discretization. The propagation along the path is obtained by iterating U such that the entire evolution along the L-step discretization is generated by U^L. Let $U_1, \ldots, U_L$ be arbitrary unitary operators on $\mathcal{H}_{S'}$ and define $U :=$

$\sum_{l=0}^{L-1} U_l \otimes |c_{l+1}\rangle\langle c_l| + U_L \otimes |c_0\rangle\langle c_L|$. One can confirm that U is unitary. The unitary operator $V(U)$ is energy conserving, and analogous to section G 3 a one can confirm that $V(\hat{1}_{S'} \otimes |c_l\rangle\langle c_l|) = \hat{1}_{S'} \otimes |c_l\rangle\langle c_l| \otimes \hat{1}_E$, as well as $V(U)^L[\hat{1}_{S'} \otimes |c_i\rangle\langle c_i| \otimes \hat{1}_E] = [\hat{1}_{S'} \otimes |c_f\rangle\langle c_f| \otimes \hat{1}_E]V(U)^L$. Hence, $V := V(U)^L$ satisfies the conditions in Assumptions 1.

c. A minimal example with time-reversal

Here we consider a setup that satisfies the conditions in Assumptions 2. With $H^i_S, H^f_{S'}$ as in section G 3 a, and H_E the energy-ladder, let $\{|c_{i+}\rangle, |c_{i-}\rangle, |c_{f+}\rangle, |c_{f-}\rangle\}$ be orthonormal elements spanning the Hilbert space $\mathcal{H}_C$ of the control system C. Let

$$\begin{aligned} H_{S'C} :=& H^i_{S'} \otimes P^i_C + H^f_{S'} \otimes P^f_C, \\ P^i_C :=& |c_{i+}\rangle\langle c_{i+}| + |c_{i-}\rangle\langle c_{i-}|, \\ P^f_C :=& |c_{f+}\rangle\langle c_{f+}| + |c_{f-}\rangle\langle c_{f-}|, \\ H :=& H_{S'C} \otimes \hat{1}_E + \hat{1}_{SC} \otimes H_E. \end{aligned} \tag{G16}$$

We next turn to the time-reversals. On $\mathcal{H}_C$ we define

$$\begin{aligned} Y :=& |c_{i+}\rangle\langle c_{i-}| + |c_{i-}\rangle\langle c_{i+}| \\ &+ |c_{f+}\rangle\langle c_{f-}| + |c_{f-}\rangle\langle c_{f+}|. \end{aligned} \tag{G17}$$

As one can see, Y is a unitary operator on $\mathcal{H}_C$. Define the basis

$$\begin{aligned} \mathcal{B} :=& \{|\chi^i_n\rangle|c_{i+}\rangle\}_n \cup \{|\chi^i_n\rangle|c_{i-}\rangle\}_n \\ &\cup \{|\chi^f_n\rangle|c_{f+}\rangle\}_n \cup \{|\chi^f_n\rangle|c_{f-}\rangle\}_n, \end{aligned} \tag{G18}$$

where $\{|\chi^i_n\rangle\}_n$ is an orthonormal eigenbasis of $H^i_{S'}$ and $\{|\chi^f_n\rangle\}_n$ is an orthonormal eigenbasis of $H^f_{S'}$. Define $\mathcal{T}_{S'C}(Q) := [\hat{1}_{S'} \otimes Y]Q^t[\hat{1}_{S'} \otimes Y^\dagger]$, where t denotes the transpose with respect to the basis $\mathcal{B}$. Note that $\hat{1}_{S'} \otimes Y$ is complex symmetric with respect to the basis $\mathcal{B}$ (and the space $\mathcal{H}_{S'C}$ on which it operates is finite-dimensional). Hence, according to Proposition 2 it follows that $\mathcal{T}_{S'C}$ is a time-reversal (and is moreover the transpose with respect to some basis). One can verify that $\mathcal{T}_{S'C}(H_{S'C}) = H_{S'C}$, and furthermore

$$\begin{aligned} \mathcal{T}_{S'C}(\hat{1}_{S'} \otimes |c_{i+}\rangle\langle c_{i+}|) =& \hat{1}_{S'} \otimes |c_{i-}\rangle\langle c_{i-}|, \\ \mathcal{T}_{S'C}(\hat{1}_{S'} \otimes |c_{f+}\rangle\langle c_{f+}|) =& \hat{1}_{S'} \otimes |c_{f-}\rangle\langle c_{f-}|. \end{aligned} \tag{G19}$$

We moreover define the time-reversal $\mathcal{T}_E$ on the energy reservoir as the transpose with respect to the basis $\{|j\rangle\}_{j\in\mathbb{Z}}$, and thus it is the case that $\mathcal{T}_E(H_E) = H_E$. We define the global time-reversal as $\mathcal{T} := \mathcal{T}_{S'C} \otimes \mathcal{T}_E$. Let U^+ and $\overline{U}^+$ be arbitrary unitary operators on $\mathcal{H}_{S'}$, and define

$$\begin{aligned} U :=& U^+ \otimes |c_{f+}\rangle\langle c_{i+}| + \overline{U}^+ \otimes |c_{i+}\rangle\langle c_{f+}| \\ &+ U^- \otimes |c_{i-}\rangle\langle c_{f-}| + \overline{U}^- \otimes |c_{f-}\rangle\langle c_{i-}| \end{aligned} \tag{G20}$$

where $U^- := \sum_{nn'} |\chi^i_n\rangle\langle\chi^f_{n'}|U^+|\chi^i_n\rangle\langle\chi^f_{n'}|$ and $\overline{U}^- := \sum_{nn'} |\chi^f_n\rangle\langle\chi^i_{n'}|\overline{U}^+|\chi^f_n\rangle\langle\chi^i_{n'}|$. One can confirm that U^-, $\overline{U}^-$, and U are unitary, and thus $V(U)$ is an energy conserving unitary operator. One can moreover confirm that $\mathcal{T}_{S'C}(U) = U$. Since $\mathcal{T}_{S'C}(H_{S'C}) = H_{S'C}$ (and since $\mathcal{H}_{S'C}$ is finite-dimensional) we know by Lemma 13 that $\mathcal{T}(V(U)) = V(\mathcal{T}_{S'C}(U)) = V(U)$. Hence, the dynamics is time-reversal symmetric. With a reasoning analogous to section G 3 a, one can also show that $V(U)[\hat{1}_{S'} \otimes |c_{i+}\rangle\langle c_{i+}| \otimes \hat{1}_E] = [\hat{1}_{S'} \otimes |c_{f+}\rangle\langle c_{f+}| \otimes \hat{1}_E]V(U)$. Hence, all the conditions of Assumptions 2 are satisfied.

d. Discretized paths of Hamiltonians with time-reversal

Here we modify the setup of section G 3 b such that it incorporates time-reversal, and satisfies the conditions in Assumptions 2.

We let $\{|c^\pm_l\rangle\}_{l=0}^L$ be a set of orthonormal elements spanning the Hilbert space $\mathcal{H}_C$ of the control system C. For the family of Hermitian operators $(\widetilde{H}_l)_{l=0}^L$, let

$$\begin{aligned} H_{S'C} :=& \sum_{l=0}^{L} \widetilde{H}_l \otimes P^l_C, \\ P^l_C :=& |c^+_l\rangle\langle c^+_l| + |c^-_l\rangle\langle c^-_l|, \\ H :=& H_{S'C} \otimes \hat{1}_E + \hat{1}_{SC} \otimes H_E. \end{aligned} \tag{G21}$$

To compare with Assumptions 2 we have

$$\begin{aligned} &|c_{i\pm}\rangle = |c^\pm_0\rangle, \quad |c_{f\pm}\rangle = |c^\pm_L\rangle, \\ &H^i_{S'} = \widetilde{H}_0, \quad H^f_{S'} = \widetilde{H}_L, \\ &P^i_C = P^0_C, \quad P^f_C = P^L_C, \\ &H^\perp = \sum_{l=1}^{L-1} \widetilde{H}_l \otimes P^l_C. \end{aligned}$$

We next turn to the time-reversals. On $\mathcal{H}_C$ we define $Y := \sum_{l=0}^L (|c^+_l\rangle\langle c^-_l| + |c^-_l\rangle\langle c^+_l|)$. One can confirm that Y is a unitary operator. Note that $\mathcal{B} := \{|\chi^{(l)}_n\rangle|c^+_l\rangle\}_{l,n} \cup \{|\chi^{(l)}_n\rangle|c^-_l\rangle\}_{l,n}$ is an orthonormal basis of $\mathcal{H}_{S'C}$. We define the time-reversal $\mathcal{T}_{S'C}(Q) := [\hat{1}_{S'} \otimes Y]Q^t[\hat{1}_{S'} \otimes Y^\dagger]$, where t denotes the transpose with respect to the basis $\mathcal{B}$. One can show that $\mathcal{T}_{S'C}(H_{S'C}) = H_{S'C}$ and

$$\begin{aligned} \mathcal{T}_{S'C}(\hat{1}_{S'} \otimes |c_{i+}\rangle\langle c_{i+}|) =& \hat{1}_{S'} \otimes |c_{i-}\rangle\langle c_{i-}|, \\ \mathcal{T}_{S'C}(\hat{1}_{S'} \otimes |c_{f+}\rangle\langle c_{f+}|) =& \hat{1}_{S'} \otimes |c_{f-}\rangle\langle c_{f-}|. \end{aligned}$$

The time-reversal $\mathcal{T}_E$ is defined as the transpose with respect to the basis $\{|j\rangle\}_{j\in\mathbb{Z}}$ of the energy-ladder. Let $U^+_0, \ldots, U^+_L$ be arbitrary unitary operators on $\mathcal{H}_{S'}$, and define

$$\begin{aligned} U^-_L :=& \sum_{nn'} |\chi^{(L)}_{n'}\rangle\langle\chi^{(0)}_n|U^+_L|\chi^{(L)}_{n'}\rangle\langle\chi^{(0)}_n|, \\ U^-_s :=& \sum_{nn'} |\chi^{(s)}_{n'}\rangle\langle\chi^{(s+1)}_n|U^+_s|\chi^{(s)}_{n'}\rangle\langle\chi^{(s+1)}_n|, \ s = 0, \ldots, L-1, \end{aligned}$$

as well as the unitary operator

$$U := U_L^+ \otimes |c_0^+\rangle\langle c_L^+| + \sum_{s=0}^{L-1} U_s^+ \otimes |c_{s+1}^+\rangle\langle c_s^+| + U_L^- \otimes |c_L^-\rangle\langle c_0^-| + \sum_{s=0}^{L-1} U_s^- \otimes |c_s^-\rangle\langle c_{s+1}^-|.$$

One can confirm that U is invariant with respect to $\mathcal{T}_{S'C}$. By an analogous reasoning as in section G 3 c one can show that the energy conserving unitary operator $V(U)$ satisfies $\mathcal{T}(V(U)) = V(U)$. By the general properties of time-reversals it thus follows that the operator $V := V(U)^L$ is an energy conserving, time-reversal symmetric, and unitary operator. The proof that V also satisfies perfect control $V[\hat{1}_{S'} \otimes |c_{i+}\rangle\langle c_{i+}| \otimes \hat{1}_E] = [\hat{1}_{S'} \otimes |c_{i+}\rangle\langle c_{i+}| \otimes \hat{1}_E]V$ can be done analogously as to section G 3 b, and thus we can conclude that $V := V(U)^L$ satisfies all the conditions of Assumptions 2.

Appendix H: Conditional fluctuation relations

Here we consider a generalized type of fluctuation relations that naturally includes non-equilibrium states. This extension may at first sight seem rather radical. However, our quantum fluctuation relation in Proposition 3 strictly speaking already requires initial non-equilibrium states, due to the control system, as discussed in section A 3.

As we have seen in section C 4, the assumption of perfect control is a rather strong condition, and typically requires an energy reservoir spectrum that is unbounded from below (as well as above). The conditional fluctuation relations allow us to abolish the perfect control (see section H 7 a). Not only can we avoid unbounded spectra, but we can also base the conditional fluctuation relations on finite-dimensional Hilbert spaces (see section H 7 b for an explicit example).

1. The Gibbs map and the partition map

For a given operator A we define the Gibbs map $\mathcal{G}_A$ and the partition map $\mathcal{Z}_A$ by

$$\mathcal{G}_A(Q) := \frac{1}{\mathcal{Z}_A(Q)}\mathcal{J}_A(Q), \quad \mathcal{Z}_A(Q) := \mathrm{Tr}\mathcal{J}_A(Q).$$

By construction, $\mathcal{G}_A(Q)$ is a density operator whenever Q is a positive operator (modulo the existence of $\mathcal{Z}_A(Q)$). In the special case that $A = \beta H$ for $\beta \geq 0$, then $\mathcal{Z}_{\beta H}(\hat{1}) = Z_\beta(H)$ and $\mathcal{G}_{\beta H}(\hat{1}) = G_\beta(H)$.

An immediate question is what class of density operators that can be reached by the Gibbs map. If H is a bounded Hermitian operator, then $e^{\pm\beta H}$ is a also bounded, and thus $\|\mathcal{J}_{-\beta H}(\rho)\| < +\infty$, where $\|Q\| := \sup_{\|\psi\|} \|Q|\psi\rangle\|$ denotes the standard operator norm. For an arbitrary density operator ρ (which by virtue of being trace class also is bounded, see e.g. [90]) let $Q := \mathcal{J}_{-\beta H}(\rho)/\|\mathcal{J}_{-\beta H}(\rho)\|$. By construction $0 \leq Q \leq 1$, and one can confirm that that $\mathcal{G}_{\beta H}(Q) = \rho$. Hence, for bounded Hermitian operators H, one can reach all density operators via the Gibbs map (and thus in particular if the Hilbert space is finite-dimensional). The issue becomes more complicated if H is unbounded, since we have to take into account the domain of definition of H and of $e^{\pm H}$. More generally it may be the case that the Gibbs map does not generate the entire set of density operators. Although we do use the Gibbs map for unbounded H, we will nevertheless not consider this question further in this investigation.

One can further note that $\mathcal{G}_{\beta H}$ is a many-to-one map from the set of POVM elements, although in a relatively mild sense. If Q is a POVM element, then rQ for $0 < r < 1$ is also a valid POVM element. The Gibbs map $\mathcal{G}_{\beta H}$ maps both Q and rQ to the same density operator.

2. Without time-reversal

As mentioned above, the fact that we drop the assumption of perfect control implies a simpler structure. The first simplification is that we here only need to consider two subsystems: the energy reservoir E and the rest $\widetilde{S}$. For the general theory there is no need for any further partitioning into subsystems, but to relate to the results in previous sections we would let $\widetilde{S} = S'C = SBC$.

Assumptions 3. *Let $\mathcal{H}_{\widetilde{S}}$ and $\mathcal{H}_E$ be complex Hilbert spaces*

- *Let $H_{\widetilde{S}}$ and H_E be Hermitian operators on $\mathcal{H}_{\widetilde{S}}$ and $\mathcal{H}_E$, respectively, and let*

$$H := H_{\widetilde{S}} \otimes \hat{1}_E + \hat{1}_{\widetilde{S}} \otimes H_E.$$

- *Let V be a unitary operator on $\mathcal{H}_{\widetilde{S}} \otimes \mathcal{H}_E$ such that $[H, V] = 0$.*
- *Let $Q^i_{\widetilde{S}}$ and $Q^f_{\widetilde{S}}$ be operators on $\mathcal{H}_{\widetilde{S}}$ such that $0 \leq Q^i_{\widetilde{S}} \leq \hat{1}_{\widetilde{S}}$ and $0 \leq Q^f_{\widetilde{S}} \leq \hat{1}_{\widetilde{S}}$.*

The operators $Q^i_{\widetilde{S}}$ and $Q^f_{\widetilde{S}}$ play dual roles in this analysis. First, they correspond to control measurements. For example, we can form the two POVMs $\{Q^i_{\widetilde{S}}, \hat{1}_{\widetilde{S}} - Q^i_{\widetilde{S}}\}$ and $\{Q^f_{\widetilde{S}}, \hat{1}_{\widetilde{S}} - Q^f_{\widetilde{S}}\}$. In these POVMs $Q^i_{\widetilde{S}}$ and $Q^f_{\widetilde{S}}$ are the 'successful' outcomes, and the CPMs $\tilde{\mathcal{F}}$ and $\tilde{\mathcal{R}}$, defined in (H1) below, maps to the corresponding (non-normalized) post-measurement states of the reservoir conditioned on these successful outcomes. (Nothing in this formalism forces us to use binary POVMs. See the discussion at the end of section H 3.)

The second role of $Q^i_{\widetilde{S}}$ and $Q^f_{\widetilde{S}}$ is that the they parametrize initial states via the Gibbs map $\mathcal{G}_{\beta H_{\widetilde{S}}}$. These

roles are switched within the pair, such that for the forward process $Q^i_{\widetilde{S}}$ gives the initial state, and $Q^f_{\widetilde{S}}$ defines the control measurement. For the reverse process, $Q^f_{\widetilde{S}}$ gives the initial state, and $Q^i_{\widetilde{S}}$ the measurement.

We define the completely positive maps

$$\begin{aligned}\tilde{\mathcal{F}}(\sigma) =& \mathrm{Tr}_{\widetilde{S}}([Q^f_{\widetilde{S}} \otimes \hat{1}_E]V[\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}}) \otimes \sigma]V^\dagger),\\ \tilde{\mathcal{R}}(\sigma) =& \mathrm{Tr}_{\widetilde{S}}([Q^i_{\widetilde{S}} \otimes \hat{1}_E]V^\dagger[\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}}) \otimes \sigma]V).\end{aligned} \tag{H1}$$

The following is the counterpart of Lemma 1, and is left without proof.

Lemma 16. *With V, $H_{\widetilde{S}}$, and H_E as in Assumptions 3, it is the case that for every $\alpha \in \mathbb{C}$*

$$\begin{aligned}V[e^{\alpha H_{\widetilde{S}}} \otimes \hat{1}_E] &= [e^{\alpha H_{\widetilde{S}}} \otimes e^{\alpha H_E}]V[\hat{1}_{\widetilde{S}} \otimes e^{-\alpha H_E}],\\ [e^{\alpha H_{\widetilde{S}}} \otimes \hat{1}_E]V &= [\hat{1}_{\widetilde{S}} \otimes e^{-\alpha H_E}]V[e^{\alpha H_{\widetilde{S}}} \otimes e^{\alpha H_E}].\end{aligned} \tag{H2}$$

Proposition 8. *With Assumptions 3, the CPMs $\tilde{\mathcal{F}}$ and $\tilde{\mathcal{R}}$ as defined in (H1) satisfy*

$$\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}})\tilde{\mathcal{F}} = \mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}})\mathcal{J}_{\beta H_E}\tilde{\mathcal{R}}^*\mathcal{J}^{-1}_{\beta H_E}. \tag{H3}$$

Proof. By comparing the definition of the CPM $\tilde{\mathcal{R}}$ with Eqs. (A6) and (A7) in section A 5, we can conclude that

$$\begin{aligned}&\tilde{\mathcal{R}}^*(Y)\\ &= \mathrm{Tr}_{\widetilde{S}}([\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}}) \otimes \hat{1}_E]V[Q^i_{\widetilde{S}} \otimes Y]V^\dagger)\\ &= \frac{1}{\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}})}\mathrm{Tr}_{\widetilde{S}}\Big([Q^f_{\widetilde{S}} \otimes \hat{1}_E]\\ &\quad [e^{-\beta H_{\widetilde{S}}/2} \otimes \hat{1}_E]V[Q^i_{\widetilde{S}} \otimes Y]V^\dagger[e^{-\beta H_{\widetilde{S}}/2} \otimes \hat{1}_E]\Big)\\ &\text{[Lemma 16]}\\ &= \frac{1}{\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}})}e^{\beta H_E/2}\mathrm{Tr}_{\widetilde{S}}\Big([Q^f_{\widetilde{S}} \otimes \hat{1}_E]V\\ &\quad [e^{-\beta H_{\widetilde{S}}/2}Q^i_{\widetilde{S}}e^{-\beta H_{\widetilde{S}}/2} \otimes \mathcal{J}_{\beta H_E}(Y)]V^\dagger\Big)e^{\beta H_E/2}\\ &= \frac{\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}})}{\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}})}\mathcal{J}^{-1}_{\beta H_E} \circ \tilde{\mathcal{F}}_+ \circ \mathcal{J}_{\beta H_E}(Y).\end{aligned} \tag{H4}$$

Hence $\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}})\tilde{\mathcal{R}}^* = \mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}})\mathcal{J}^{-1}_{\beta H_E} \circ \tilde{\mathcal{F}}_+ \circ \mathcal{J}_{\beta H_E}$. By multiplying from the left with $\mathcal{J}_{\beta H_E}$ and from the right with $\mathcal{J}^{-1}_{\beta H_E}$ we obtain Eq. (H3). □

3. With time-reversal

Assumptions 4. *Let $\mathcal{H}_{\widetilde{S}}$ and $\mathcal{H}_E$ be complex Hilbert spaces. Let $\mathcal{T}_{\widetilde{S}}$ and $\mathcal{T}_E$ be time-reversals on $\widetilde{S}$ and E, respectively, and let $\mathcal{T} := \mathcal{T}_{\widetilde{S}} \otimes \mathcal{T}_E$.*

- *Let $H_{\widetilde{S}}$ and H_E be Hermitian operators on $\mathcal{H}_{\widetilde{S}}$ and $\mathcal{H}_E$, respectively, and let*

$$H := H_{\widetilde{S}} \otimes \hat{1}_E + \hat{1}_{\widetilde{S}} \otimes H_E. \tag{H5}$$

- *Let V be a unitary operator on $\mathcal{H}_{\widetilde{S}} \otimes \mathcal{H}_E$ such that $[H, V] = 0$.*
- *Let $Q^{i+}_{\widetilde{S}}$ and $Q^{f+}_{\widetilde{S}}$ be an operator on $\mathcal{H}_{\widetilde{S}}$ such that $0 \le Q^{i+}_{\widetilde{S}} \le \hat{1}_{\widetilde{S}}$ and $0 \le Q^{f+}_{\widetilde{S}} \le \hat{1}_{\widetilde{S}}$.*
- *Let $\mathcal{T}_E(H_E) = H_E$, $\mathcal{T}_{\widetilde{S}}(H_{\widetilde{S}}) = H_{\widetilde{S}}$, and $\mathcal{T}(V) = V$. Define $Q^{i-}_{\widetilde{S}} := \mathcal{T}_{\widetilde{S}}(Q^{i+}_{\widetilde{S}})$ and $Q^{f-}_{\widetilde{S}} := \mathcal{T}_{\widetilde{S}}(Q^{f+}_{\widetilde{S}})$.*

These assumptions are constructed such that the triple $V, Q^i_{\widetilde{S}} := Q^{i+}_{\widetilde{S}}, Q^f_{\widetilde{S}} := Q^{f+}_{\widetilde{S}}$ satisfies Assumptions 3. Simultaneously, the triple $V, Q^i_{\widetilde{S}} := Q^{f-}_{\widetilde{S}}, Q^f_{\widetilde{S}} := Q^{i-}_{\widetilde{S}}$ also satisfies Assumptions 3.

By Lemma 3 and Lemma 4 it follows that $0 \le Q^{i+}_{\widetilde{S}} \le \hat{1}_{\widetilde{S}}$ implies $0 \le Q^{i-}_{\widetilde{S}} \le \hat{1}_{\widetilde{S}}$ and analogous for $Q^{f-}_{\widetilde{S}}$.

Lemma 17. *Let $\mathcal{T}$ be a time-reversal and A an operator such that $\mathcal{T}(A) = A$. Then*

$$\begin{aligned}&\mathcal{J}_A\mathcal{T} = \mathcal{T}\mathcal{J}_A,\\ &\mathcal{Z}_A\big(\mathcal{T}(Q)\big) = \mathcal{Z}_A(Q),\\ &\mathcal{G}_A\big(\mathcal{T}(Q)\big) = \mathcal{T}\big(\mathcal{G}_A(Q)\big),\\ &\mathrm{Tr}\big(Q\mathcal{J}_A(R)\big) = \mathrm{Tr}\big(\mathcal{J}_{A^\dagger}(Q)R\big).\end{aligned} \tag{H6}$$

Hence if $A^\dagger = A$, then $\mathrm{Tr}\big(Q\mathcal{J}_A(R)\big) = \mathrm{Tr}\big(\mathcal{J}_A(Q)R\big)$.

Define the CPMs

$$\begin{aligned}\tilde{\mathcal{F}}_+(\sigma) =& \mathrm{Tr}_{\widetilde{S}}([Q^{f+}_{\widetilde{S}} \otimes \hat{1}_E]V[\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{i+}_{\widetilde{S}}) \otimes \sigma]V^\dagger),\\ \tilde{\mathcal{R}}_+(\sigma) =& \mathrm{Tr}_{\widetilde{S}}([Q^{i+}_{\widetilde{S}} \otimes \hat{1}_E]V^\dagger[\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{f+}_{\widetilde{S}}) \otimes \sigma]V),\\ \tilde{\mathcal{F}}_-(\sigma) =& \mathrm{Tr}_{\widetilde{S}}([Q^{i-}_{\widetilde{S}} \otimes \hat{1}_E]V[\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{f-}_{\widetilde{S}}) \otimes \sigma]V^\dagger),\\ \tilde{\mathcal{R}}_-(\sigma) =& \mathrm{Tr}_{\widetilde{S}}([Q^{f-}_{\widetilde{S}} \otimes \hat{1}_E]V^\dagger[\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{i-}_{\widetilde{S}}) \otimes \sigma]V).\end{aligned} \tag{H7}$$

The CPMs $\tilde{\mathcal{F}}_\pm$ and $\tilde{\mathcal{R}}_\pm$ describe the *unnormalized* mapping from the input to the output conditioned on the successful control measurement. The corresponding success probabilities are given by the traces. (In section H 6 b we shall consider these success probabilities in the special case of energy translation invariance.) Analogous to Lemma 11 one can prove the following.

Lemma 18. *With Assumptions 4, the CPMs $\tilde{\mathcal{R}}_+$ and $\tilde{\mathcal{F}}_-$ as defined in Eq. (H7) are related as*

$$\mathcal{T}_E\tilde{\mathcal{R}}_+ = \tilde{\mathcal{F}}_-\mathcal{T}_E \tag{H8}$$

and thus

$$\tilde{\mathcal{R}}^*_+ = \tilde{\mathcal{F}}^\ominus_-. \tag{H9}$$

The proof is obtained by combining the definition of $\tilde{\mathcal{R}}_+$ in Eq. (H7) with the general fact that $\mathcal{T}_1\big(\mathrm{Tr}_2(\rho)\big) = \mathrm{Tr}_2\big([\mathcal{T}_1 \otimes \mathcal{T}_2](\rho)\big)$, and the time-reversal symmetry.

Proposition 9 (Conditional quantum fluctuation relation)**.** *With Assumptions 4, the CPMs $\tilde{\mathcal{F}}_+$ and $\tilde{\mathcal{F}}_-$ as defined in Eq. (H7), are related as*

$$\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}})\tilde{\mathcal{F}}_+ = \mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}})\mathcal{J}_{\beta H_E}\tilde{\mathcal{F}}^{\ominus}_-\mathcal{J}^{-1}_{\beta H_E}. \tag{H10}$$

Here we use the notation $\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}}) := \mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^{i+}_{\widetilde{S}}) = \mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^{i-}_{\widetilde{S}})$ and $\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}}) := \mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^{f+}_{\widetilde{S}}) = \mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^{f-}_{\widetilde{S}})$.

Proof. The triple V, $Q^{i+}_{\widetilde{S}}$, and $Q^{f+}_{\widetilde{S}}$ from Assumptions 4 satisfies Assumptions 3 with $Q^i_{\widetilde{S}} := Q^{i+}_{\widetilde{S}}$, $Q^f_{\widetilde{S}} := Q^{f+}_{\widetilde{S}}$. Hence, Proposition 8 is applicable and yields $\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}})\tilde{\mathcal{F}}_+ = \mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}})\mathcal{J}_{\beta H_E}\tilde{\mathcal{R}}^*_+\mathcal{J}^{-1}_{\beta H_E}$. The application of Lemma 18 to the above equation results in equation (H10). □

Note that Proposition 9 only makes a statement concerning pairs of measurement operators. It does not make any assumptions on to POVMs that these measurement operators may be members of. For example, instead of basing the induced CPMs in (H7) on the pair $(Q^{i+}_{\widetilde{S}}, Q^{f+}_{\widetilde{S}})$ we could equally well obtain fluctuation relations for each of the pairs $(Q^{i+}_{\widetilde{S}}, \hat{1}-Q^{f+}_{\widetilde{S}})$, $(\hat{1}-Q^{i+}_{\widetilde{S}}, Q^{f+}_{\widetilde{S}})$, and $(\hat{1}-Q^{i+}_{\widetilde{S}}, \hat{1}-Q^{f+}_{\widetilde{S}})$. There is also no need to assume that the POVMs are binary. For two POVMs $\{Q^{i+}_{\widetilde{S},k}\}_k$ and $\{Q^{f+}_{\widetilde{S},l}\}_l$ one can, for each possible combination of POVM elements $(Q^{i+}_{\widetilde{S},k}, Q^{f+}_{\widetilde{S},l})$, construct the corresponding CPMs $\tilde{\mathcal{F}}^{(k,l)}_\pm$ as in (H7), where each of these pairs satisfies the conditional fluctuation relation (H10).

4. Generally no decoupling of diagonals

In section D 1 we showed that the channels $\mathcal{F}_\pm$ induced on the reservoir are such that the dynamics of the modes of coherence decouples, thus yielding a separation into several off-diagonal Crooks relations. The reader may wonder whether something similar is true for the conditional CPMs $\tilde{\mathcal{F}}_\pm$. In section D 1 the starting point was Lemma 12, which shows that $\mathcal{F}_\pm$ and $\mathcal{R}_\pm$ commute with the commutator with respect to H_E. The following Lemma shows that this is generally not true for the conditional CPMs $\tilde{\mathcal{F}}_\pm$ and $\tilde{\mathcal{R}}_\pm$, and thus we cannot expect to have a separation in to diagonal and off-diagonal fluctuation relations in this more general case. (For an explicit example of such 'mixing' of diagonal and off-diagonal elements, see section H 7 b.)

Lemma 19. *With Assumptions 4, the CPM $\tilde{\mathcal{F}}_+$ defined in Eq. (H7) satisfies the following relation*

$$\begin{aligned}[H_E, \tilde{\mathcal{F}}_+(\sigma)] &= \tilde{\mathcal{F}}_+([H_E,\sigma]) \\ &+ \mathrm{Tr}_{\widetilde{S}}\Big((Q^{f+}_{\widetilde{S}}\otimes\hat{1}_E)V\big(\mathcal{G}_{\beta H_{\widetilde{S}}}([H_{\widetilde{S}}, Q^{i+}_{\widetilde{S}}])\otimes\sigma\big)V^\dagger\Big) \\ &+ \mathrm{Tr}_{\widetilde{S}}\Big(([H_{\widetilde{S}}, Q^{f+}_{\widetilde{S}}]\otimes\hat{1}_E)V\big(\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{i+}_{\widetilde{S}})\otimes\sigma\big)V^\dagger\Big).\end{aligned} \tag{H11}$$

Analogous statements hold for $\tilde{\mathcal{F}}_-$ and $\tilde{\mathcal{R}}_\pm$.

The proof is a straightforward verification, much analogous to the reasoning in the proof of Lemma 12.

5. Nevertheless a diagonal conditional fluctuation relation

Since there is generally no decoupling between the modes of coherence, it may at first sight seem like there is no separation into diagonal and off-diagonal Crooks relations. However, it turns out that the diagonal fluctuation relation exists generally.

In section D 2 we described the evolution of the diagonal elements with respect to the basis $\{|n\rangle\}_n$ in terms of the conditional probabilities $p_\pm(m|n)$. We here define the analogous objects $\tilde{p}_\pm(m|n) := \langle m|\tilde{\mathcal{F}}_\pm(|n\rangle\langle n|)|m\rangle$. One can interpret $\tilde{p}_\pm(m|n)$ as the probability that we would measure the energy reservoir in state m in addition to obtaining a successful outcome of the control measurement, given that the reservoir initially is in state n.

With Assumptions 4 and otherwise the setup of Proposition 4 we get

$$\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}})\ \tilde{p}_+(m|n) = \mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}})e^{\beta(E_n-E_m)}\tilde{p}_-(n|m). \tag{H12}$$

The proof is more or less identical to the proof of Proposition 4, but with Proposition 9 as the starting point.

One should note that although (H12) is very similar to Proposition 4, the relation (H12) should be interpreted with care since there is no decoupling between the different diagonal modes of coherence. Hence, the probability $\langle m|\tilde{\mathcal{F}}_\pm(\sigma)|m\rangle$ to detect a specific final energy eigenstate m not only depends on the diagonal elements $\langle n|\sigma|n\rangle$ of the input state, but also on the off-diagonal elements. We can thus no longer claim that an initial energy measurement would not perturb a final energy measurement.

6. The special case $[H_{\widetilde{S}}, Q^{i\pm}_{\widetilde{S}}] = 0$ and $[H_{\widetilde{S}}, Q^{f\pm}_{\widetilde{S}}] = 0$

In the case that the measurement operators $Q^{i\pm}_{\widetilde{S}}$ and $Q^{f\pm}_{\widetilde{S}}$ commute with $H_{\widetilde{S}}$ we can regain several of the properties of the unconditional fluctuation relations. By a direct application of Lemma 19 we get $[H_E, \tilde{\mathcal{F}}_\pm(\sigma)] = \tilde{\mathcal{F}}_\pm([H_E,\sigma])$. Analogously to how we obtained Corollary 1 from Lemma 12, and with analogous assumptions, we also regain the decoupling of the different off-diagonal

modes of coherence, i.e., $\langle m'|\tilde{\mathcal{F}}_\pm(|n'\rangle\langle n|)|m\rangle = 0$, if $E_{m'} - E_{n'} \neq E_m - E_n$. It follows that much of the arguments in section G can be reiterated. Due to the decoupling it is again meaningful to define the transitions among the elements of a diagonal with a specified offset δ, $\tilde{q}^\delta_\pm(m|n) := \langle m|\tilde{\mathcal{F}}_\pm(|n\rangle\langle n'|)|m'\rangle$. As the counterpart of Proposition 5 we obtain a Crooks relation for each δ,

$$\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}})\tilde{q}^\delta_+(m|n) = e^{\beta(E_n - E_m)}\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}})\tilde{q}^\delta_-(n|m).$$

a. A classical conditional Crooks relation

By additionally assuming the energy translation invariant model as in section G one regains the energy translation invariance of the induced CPMs $\Delta^j\tilde{\mathcal{F}}_\pm(\sigma){\Delta^\dagger}^k = \tilde{\mathcal{F}}_\pm\big(\Delta^j\sigma{\Delta^\dagger}^k\big)$. Analogous to section G 1 we can also define

$$\tilde{P}_\pm(w) := \sum_{j,j':j-j'=w/s} \tilde{p}_\pm(j'|j)\langle j|\sigma|j\rangle, \tag{H13}$$

$\tilde{P}_\pm(w)$ can be interpreted as the probability that the energy reservoir looses the energy w and that the control measurement is successful. The decoupling again guarantees the stability of the transition probabilities under repeated energy measurements. We obtain the classical conditional Crooks relation

$$\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}})\tilde{P}_+(w) = e^{\beta w}\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}})\tilde{P}_-(-w) \tag{H14}$$

in a manner very similar to what we did in section G 1.

b. A classical conditional Jarzynski relation

The control measurements do not succeed with unit probability. However, for the energy translation symmetric case, with diagonal measurement operators, the success probabilities $\mathrm{Tr}\mathcal{F}_\pm(\sigma)$ and $\mathrm{Tr}\mathcal{R}_\pm(\sigma)$ become independent of the state σ of the energy reservoir. To see this, we first define the following transition probabilities that do not involve the energy reservoir.

$$\begin{aligned} f_+ :=& \mathrm{Tr}\big(Q^{f+}_{\widetilde{S}}U\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{i+}_{\widetilde{S}})U^\dagger\big),\\ r_+ :=& \mathrm{Tr}\big(Q^{i+}_{\widetilde{S}}U^\dagger\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{f+}_{\widetilde{S}})U\big),\\ f_- :=& \mathrm{Tr}\big(Q^{i-}_{\widetilde{S}}U\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{f-}_{\widetilde{S}})U^\dagger\big),\\ r_- :=& \mathrm{Tr}\big(Q^{f-}_{\widetilde{S}}U^\dagger\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{i-}_{\widetilde{S}})U\big). \end{aligned} \tag{H15}$$

In words f_+ is the probability that we would obtain the 'successful' outcome when we measure the POVM $\{Q^{f+}_{\widetilde{S}}, \hat{1}_{\widetilde{S}} - Q^{f+}_{\widetilde{S}}\}$ if the initial state $\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{i+}_{\widetilde{S}})$ is evolved under U.

One should keep in mind that since $Q^{f\pm}_{\widetilde{S}}$ and $Q^{i\pm}_{\widetilde{S}}$ commute with $H_{\widetilde{S}}$, it follows that the above expressions do not involve coherences with respect to the energy egienbases. For example, if $H_{\widetilde{S}}$ is non-degenerate with eigenstates $|\psi_n\rangle$ then $f_+ = \sum_{nn'}\langle\psi_n|Q^{f+}_{\widetilde{S}}|\psi_n\rangle|\langle\psi_n|U|\psi_{n'}\rangle|^2\langle\psi_{n'}|\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{i+}_{\widetilde{S}})|\psi_{n'}\rangle$.

One can confirm that the CPMs $\tilde{\mathcal{F}}_\pm$ and $\tilde{\mathcal{R}}_\pm$ in (H7) with $V := V(U)$ as in (G2), and $[Q^{i\pm}_{\widetilde{S}}, H_{\widetilde{S}}] = 0$, $[Q^{f\pm}_{\widetilde{S}}, H_{\widetilde{S}}] = 0$, satisfy the following relations

$$\mathrm{Tr}\mathcal{F}_\pm(\sigma) = f_\pm\mathrm{Tr}(\sigma), \quad \mathrm{Tr}\mathcal{R}_\pm(\sigma) = r_\pm\mathrm{Tr}(\sigma), \tag{H16}$$

$$\tilde{\mathcal{F}}_\pm(\hat{1}_E) = f_\pm\hat{1}_E, \quad \tilde{\mathcal{R}}_\pm(\hat{1}_E) = r_\pm\hat{1}_E. \tag{H17}$$

Hence, the success probabilities of the control measurements are independent of the state of the energy reservoir. It is instructive to write the expression for $\mathcal{F}_+$ in (H16) in full

$$\begin{aligned} &\mathrm{Tr}\big([Q^{f+}_{\widetilde{S}}\otimes\hat{1}_E]V(U)[\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{i+}_{\widetilde{S}})\otimes\sigma]V^\dagger(U)\big)\\ &= \mathrm{Tr}\big(Q^{f+}_{\widetilde{S}}U\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{i+}_{\widetilde{S}})U^\dagger\big). \end{aligned}$$

Hence, in terms of the success probability, the experiment involving the energy reservoir behaves as if it was a simpler experiment not including the reservoir, where the unitary $V(U)$ is replaced by U (but one should keep in mind that this relies on the assumptions that $Q^{f+}_{\widetilde{S}}$ and $Q^{i+}_{\widetilde{S}}$ that $Q^{i\pm}_{\widetilde{S}}$ and $Q^{f\pm}_{\widetilde{S}}$ commutes with $H_{\widetilde{S}}$, and thus block-diagonalizes with respect to the energy eigenspaces of $H_{\widetilde{S}}$).

By the classical conditional Crooks relation (H14) it follows that $\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}})\sum_w e^{-\beta w}\tilde{P}_+(w) = \mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}})\sum_w\tilde{P}(-w)$. As opposed to the unconditional case, $\sum_w\tilde{P}(-w)$ is generally not equal to 1, but equates to the success probability of the reverse process, i.e., $\sum_w\tilde{P}(-w) = \sum_w\tilde{p}_-(w/s|0) = \mathrm{Tr}\tilde{\mathcal{F}}_-(|0\rangle\langle 0|) = f_-$. Thus

$$\sum_w e^{-\beta w}\tilde{P}_+(w) = f_-\frac{\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}})}{\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}})}. \tag{H18}$$

Given that the control measurement succeeds, we can define the conditional probability for the work cost w as $P_+(w|Q^{f+}_{\widetilde{S}}) := \tilde{P}_+(w)/\sum_{w'}\tilde{P}_+(w') = \tilde{P}_+(w)/f_+$. By using this we can rewrite (H18) into the more symmetric form

$$\big\langle e^{-\beta W}\big|Q^{f+}_{\widetilde{S}}\big\rangle = \frac{f_-}{f_+}\frac{\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}})}{\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}})}, \tag{H19}$$

where we use the notation $\langle e^{-\beta W}|Q^{f+}_{\widetilde{S}}\rangle := \sum_w e^{-\beta w}P_+(w|Q^{f+}_{\widetilde{S}})$.

The conditional Jarzynski relation in (H19) does remind of the Jarzynski equality under feedback control [103, 118, 119], and it may be worthwhile to investigate this potential link further. However, we will not do so in this investigation.

7. Examples

a. Abolishing perfect control

As discussed in section (C 4) the perfect control (see equations (A1) or (C3) and (C4)) can lead to an energy reservoir that has an unbounded spectrum from both above and below. From a physical point of view a spectrum that is unbounded from below is a somewhat uncomfortable. With the conditional fluctuation relations we no longer need to assume perfectly functioning control systems. We could for example let the transition from initial to final control state fail if the energy reservoir runs out of energy, so to speak. Here we demonstrate this in the case of a harmonic oscillator as the energy reservoir.

We make use of a model that was introduced in quite a detail in section IV in the Supplemental Material of [31]. Due to this we will here only give the most brief description. The main point is that one can construct a family of unitary operators $V_+(U)$ that act identically to $V(U)$ as long as the energy in the energy reservoir is high enough. (We do not employ of the 'injection' of energy that was used in [31], but instead allow the procedure to fail.)

Apart from the Hamiltonian for the energy reservoir, $H_E^+ := s\sum_{j=0}^{+\infty} j|j\rangle\langle j|$ and the class of unitary operators $V_+(U)$, the rest of the model is as in section G 3 c. We let $H_{S'C} := H^i_{S'} \otimes P^i_C + H^f_{S'} \otimes P^f_C$, and let sz^i_n, $|\chi^i_n\rangle$ be the eigenvalues and eigenvectors of $H^i_{S'}$, and sz^f_n, $|\chi^f_n\rangle$ be the eigenvalues and eigenvectors of $H^f_{S'}$. To simplify the notation we let $\{|\psi_n\rangle\}_{n=1}^{4N}$ denote the orthonormal set $\{|\chi^i_n\rangle|c_{i+}\rangle, |\chi^i_n\rangle|c_{i-}\rangle, |\chi^f_n\rangle|c_{f+}\rangle, |\chi^f_n\rangle|c_{f-}\rangle\}_{n=1}^{N}$. Similarly, let z_n denote the combined set of numbers $\{z^i_n, z^f_n\}_n$. Furthermore define $z_{\max} = \max_n z_n$ and $z_{\min} = \min_n z_n$. The projectors $\{P_+^{(l)}\}_{l\geq z_{\min}}$ onto the eigenspaces of $H_{S'C} \otimes \hat{1}_E + \hat{1}_{S'C} \otimes H_E$ are

$$P_+^{(l)} = \sum_{n:l\geq z_n} |\psi_n\rangle\langle\psi_n| \otimes |l - z_n\rangle\langle l - z_n|, \quad \forall l \geq z_{\min}.$$

For all $l \geq z_{\max}$ this simplifies as $P_+^{(l)} = \sum_{n=1\ldots N} |\psi_n\rangle\langle\psi_n| \otimes |l - z_n\rangle\langle l - z_n|$. Let us define the following map

$$V_+(U) := \sum_{l\geq z_{\max}} V_l(U) + \sum_{l=z_{\min}}^{z_{\max}-1} P_+^{(l)},$$
$$V_l(U) := \sum_{n,n'} |\psi_n\rangle\langle\psi_n|U|\psi_{n'}\rangle\langle\psi_{n'}| \otimes |l-z_n\rangle\langle l-z_{n'}|.$$

For each unitary operator U on $\mathcal{H}_{S'C}$ it follows that $V_+(U)$ is unitary on $\mathcal{H}_{S'CE}$. We let $\mathcal{T}_E$ be the transpose with respect to $\{|j\rangle\}_{j\geq 0}$, and let $\mathcal{T}_{\widetilde{S}}$ be such that $\mathcal{T}_{S'C}(H_{S'C}) = H_{S'C}$. (Note Lemma 7.) From this it follows that $\mathcal{T} := \mathcal{T}_{S'C} \otimes \mathcal{T}_E$ is such that $\mathcal{T}\big(V_+(U)\big) = V_+\big(\mathcal{T}_{S'C}(U)\big)$.

One can also confirm that $V_+(U)|\psi\rangle|j\rangle = V(U)|\psi\rangle|j\rangle$ for all $|\psi\rangle \in \mathcal{H}_{S'C}$ and all $j \geq z_{\max} - z_{\min}$, where $V(U)$ is defined in G2.

More generally, in terms of the eigenprojectors $P_+^{(l)}$ of the total Hamiltonian H, it is the case that $V_+(U)P_+^{(l)} = V(U)P_+^{(l)}$ for $l \geq z_{\max}$, while $V_+(U)P_+^{(l)} = P_+^{(l)}$ for $z_{\max} - 1 \geq l \geq z_{\min}$. One can say that V_+ 'censors' our choice of U in the sense that if U entails an energy change that we cannot afford, then $V_+(U)$ avoids to perform the too expensive parts of the operation.

Define the projector $P^{\geq z_{\max}-z_{\min}} := \sum_{j\geq z_{\max}-z_{\min}} |j\rangle\langle j|$. The condition $P^{\geq z_{\max}-z_{\min}}\sigma P^{\geq z_{\max}-z_{\min}} = \sigma$ guarantees that the action of $V_+(U)$ and $V(U)$ are identical.

As an example, let us modify the setup in section G 3 c such that we replace $V(U)$ with $V_+(U)$. For the sake of illustration we consider an extreme case where $H^i_{S'} := s|\chi^i_1\rangle\langle\chi^i_1| + 2s|\chi^i_2\rangle\langle\chi^i_2|$ and $H^f_{S'} := 3s|\chi^f_1\rangle\langle\chi^f_1| + 4s|\chi^f_2\rangle\langle\chi^f_2|$. For these choices, all transitions from the initial to the final Hamiltonian requires energy. Moreover, $z_{\max} = 4$ and $z_{\min} = 1$. One furthermore finds that

$$\begin{aligned}
P_+^{(1)} =& |\chi^i_1\rangle\langle\chi^i_1| \otimes P^i_C \otimes |0\rangle\langle 0|,\\
P_+^{(2)} =& |\chi^i_1\rangle\langle\chi^i_1| \otimes P^i_C \otimes |1\rangle\langle 1| + |\chi^i_2\rangle\langle\chi^i_2| \otimes P^i_C \otimes |0\rangle\langle 0|,\\
P_+^{(3)} =& |\chi^i_1\rangle\langle\chi^i_1| \otimes P^i_C \otimes |2\rangle\langle 2| + |\chi^i_2\rangle\langle\chi^i_2| \otimes P^i_C \otimes |1\rangle\langle 1|\\
&+ |\chi^f_1\rangle\langle\chi^f_1| \otimes P^f_C \otimes |0\rangle\langle 0|,\\
P_+^{(l)} =& |\chi^i_1\rangle\langle\chi^i_1| \otimes P^i_C \otimes |l-1\rangle\langle l-1|\\
&+ |\chi^i_2\rangle\langle\chi^i_2| \otimes P^i_C \otimes |l-2\rangle\langle l-2|\\
&+ |\chi^f_1\rangle\langle\chi^f_1| \otimes P^f_C \otimes |l-3\rangle\langle l-3|\\
&+ |\chi^f_2\rangle\langle\chi^f_2| \otimes P^f_C \otimes |l-4\rangle\langle l-4|,
\end{aligned}$$

for $l \geq 4$. Hence, assuming the control in state $|c_{i+}\rangle$ and the reservoir in the vacuum state $|0\rangle$, we find that $V_+(U)[\rho \otimes |c_{i+}\rangle\langle c_{i+}| \otimes |0\rangle\langle 0|]V_+(U)^\dagger = \rho \otimes |c_{i+}\rangle\langle c_{i+}| \otimes |0\rangle\langle 0|$ no matter what U we feed into it, and thus the control measurement will always signal 'fail'. On the other hand, if there are four or more quanta of energy in the energy reservoir, then $V_+(U)[\rho\otimes|c_{i+}\rangle\langle c_{i+}|\otimes\sigma]V_+(U)^\dagger = V(U)[\rho\otimes|c_{i+}\rangle\langle c_{i+}|\otimes\sigma]V(U)^\dagger$. In particular, if we would could choose U as in (G20), then the control measurement would aways succeed.

b. $\widetilde{S}$ and E as single qubits

The conditional fluctuation relations allow us to treat energy reservoirs with finite-dimensional Hilbert spaces. Here we consider the extreme case where both $\widetilde{S}$ and the energy reservoir E are single spin-half particles.

We assume that the spins are associated with magnetic moments, and that they are affected by a constant external magnetic field, such that they are in resonance, i.e., the splitting of the eigenenergies are identical. More

precisely,

$$H_{\widetilde{S}} = H_E = -\frac{1}{2}s|0\rangle\langle 0| + \frac{1}{2}s|1\rangle\langle 1|,$$

for some $s > 0$. The identical energy gap implies that there exist non-trivial energy conserving unitary operations with respect to $H := H_{\widetilde{S}} \otimes \hat{1}_E + \hat{1}_{\widetilde{S}} \otimes H_E$. More specifically, H has the eigenenergies $-s, 0, s$ and corresponding energy eigenspaces $\mathrm{Sp}\{|0,0\rangle\}$, $\mathrm{Sp}\{|0,1\rangle, |1,0\rangle\}$, and $\mathrm{Sp}\{|1,1\rangle\}$. An energy conserving unitary operator thus has to be block diagonal with respect to these energy eigenspaces,

$$\begin{aligned} V =& e^{i\chi_-}|0\rangle\langle 0| \otimes |0\rangle\langle 0| + e^{i\chi_+}|1\rangle\langle 1| \otimes |1\rangle\langle 1| \\ &+ U_{1,1}|0\rangle\langle 0| \otimes |1\rangle\langle 1| + U_{1,2}|0\rangle\langle 1| \otimes |1\rangle\langle 0| \\ &+ U_{2,1}|1\rangle\langle 0| \otimes |0\rangle\langle 1| + U_{2,2}|1\rangle\langle 1| \otimes |0\rangle\langle 0|, \end{aligned}$$

where $\chi_+, \chi_- \in \mathbb{R}$, and where $U = [U_{j,k}]_{j,k=1,2}$ is a unitary 2×2 matrix.

Let us choose $\mathcal{T}_{\widetilde{S}}$ and $\mathcal{T}_E$ as the transposes with respect to the eigenbasis $\{|0\rangle, |1\rangle\}$ on the two spaces, respectively. Thus $\mathcal{T}_{\widetilde{S}}(H_{\widetilde{S}}) = H_{\widetilde{S}}$ and $\mathcal{T}_E(H_E) = H_E$. Let $\mathcal{T} := \mathcal{T}_{\widetilde{S}} \otimes \mathcal{T}_E$. One can confirm that $\mathcal{T}(V) = V$ if and only if

$$U = e^{i\chi} \begin{bmatrix} -e^{-i\delta}\cos\theta & \sin\theta \\ \sin\theta & e^{i\delta}\cos\theta \end{bmatrix},$$

where $\chi, \delta, \theta \in \mathbb{R}$. (There are no restrictions on χ_+, χ_-.)

The expansion of the CPMs $\tilde{\mathcal{F}}_\pm$ for arbitrary $Q^{i\pm}_{\widetilde{S}}$ and $Q^{f\pm}_{\widetilde{S}}$ in terms of the $\{|0\rangle, |1\rangle\}$ basis results in remarkably bulky and unilluminating expressions. Therefore we shall here only consider the simpler special case where $Q^{i+}_{\widetilde{S}} := \hat{1}_{\widetilde{S}}$, $Q^{f+}_{\widetilde{S}} := |\psi\rangle\langle\psi|$, $|\psi\rangle := \frac{1}{\sqrt{2}}|0\rangle + i\frac{1}{\sqrt{2}}|1\rangle$. Consequently $Q^{i-}_{\widetilde{S}} = \mathcal{T}_{\widetilde{S}}(Q^{i+}_{\widetilde{S}}) = \hat{1}_{\widetilde{S}}$, $Q^{f-}_{\widetilde{S}} = \mathcal{T}_{\widetilde{S}}(Q^{f+}_{\widetilde{S}}) = |\psi^*\rangle\langle\psi^*|$, $|\psi^*\rangle = \frac{1}{\sqrt{2}}|0\rangle - i\frac{1}{\sqrt{2}}|1\rangle$. The partition maps take the values $\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^i_{\widetilde{S}}) = e^{\beta s/2} + e^{-\beta s/2}$ and $\mathcal{Z}_{\beta H_{\widetilde{S}}}(Q^f_{\widetilde{S}}) = (e^{\beta s/2} + e^{-\beta s/2})/2$.

The Gibbs map applied to $Q^{i+}_{\widetilde{S}} = \hat{1}_{\widetilde{S}}$ gives the the initial state of the forward process

$$\mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{i+}_{\widetilde{S}}) = \frac{e^{\beta s/2}|0\rangle\langle 0| + e^{-\beta s/2}|1\rangle\langle 1|}{e^{\beta s/2} + e^{-\beta s/2}},$$

which is the Gibbs state of $H_{\widetilde{S}}$. The initial state of the reversed process is

$$\begin{aligned} \mathcal{G}_{\beta H_{\widetilde{S}}}(Q^{f-}_{\widetilde{S}}) =& \frac{1}{e^{\beta s/2} + e^{-\beta s/2}}\Big(e^{\beta s/2}|0\rangle\langle 0| \\ &+ e^{-\beta s/2}|1\rangle\langle 1| + i|0\rangle\langle 1| - i|1\rangle\langle 0|\Big). \end{aligned}$$

For the unitary operator V we assume that $\chi = \chi_\pm = 0$, $\delta = 0$ (while we let θ be arbitrary). This results in

$$\begin{aligned} \tilde{\mathcal{F}}_+(\sigma) =& \frac{1}{e^{\beta s/2} + e^{-\beta s/2}}(V_{0+}\sigma V_{0+}^\dagger + V_{1+}\sigma V_{1+}^\dagger), \\ V_{0+} :=& \frac{e^{\beta s/4}}{\sqrt{2}}(|0\rangle\langle 0| - \cos\theta|1\rangle\langle 1| - i\sin\theta|0\rangle\langle 1|), \\ V_{1+} :=& \frac{e^{-\beta s/4}}{\sqrt{2}}(|1\rangle\langle 1| + \cos\theta|0\rangle\langle 0| + i\sin\theta|1\rangle\langle 0|), \end{aligned}$$

and

$$\begin{aligned} \tilde{\mathcal{F}}_-(\sigma) =& \frac{1}{e^{\beta s/2} + e^{-\beta s/2}}(V_{0-}\sigma V_{0-}^\dagger + V_{1-}\sigma V_{1-}^\dagger), \\ V_{0-} :=& e^{\beta s/4}(|0\rangle\langle 0| - \cos\theta|1\rangle\langle 1|) \\ &- i\sin\theta e^{-\beta s/4}|1\rangle\langle 0|, \\ V_{1-} :=& e^{-\beta s/4}(|1\rangle\langle 1| + \cos\theta|0\rangle\langle 0|) \\ &+ i\sin\theta e^{\beta s/4}|0\rangle\langle 1|. \end{aligned}$$

By a slightly tedious but straightforward calculation one can confirm that $\tilde{\mathcal{F}}_\pm$ satisfy the conditional fluctuation relation (H10), as we already know that they should, due to Proposition 9.

One can also confirm that $\tilde{\mathcal{F}}_\pm$ provide examples for the fact that the conditional maps in general do not decouple the evolution of diagonal and off-diagonal elements. For example

$$\begin{aligned} \tilde{\mathcal{F}}_+(|0\rangle\langle 0|) =& \frac{1}{2}\frac{1}{e^{\beta s/2} + e^{-\beta s/2}}\Big[\sin^2\theta e^{-\beta s/2}|1\rangle\langle 1| \\ &+ (\cos^2\theta e^{-\beta s/2} + e^{\beta s/2})|0\rangle\langle 0| \\ &+ i\sin\theta\cos\theta e^{-\beta s/2}(|1\rangle\langle 0| - |0\rangle\langle 1|)\Big], \end{aligned}$$

$$\begin{aligned} \tilde{\mathcal{F}}_-(|0\rangle\langle 1|) =& \frac{\cos\theta}{e^{\beta s/2} + e^{-\beta s/2}}\Big[(e^{-\beta s/2} - e^{\beta s/2})|0\rangle\langle 1| \\ &+ i\sin\theta(|1\rangle\langle 1| - |0\rangle\langle 0|)\Big]. \end{aligned}$$

Hence, diagonal and off-diagonal elements get mixed. In particular, a diagonal state such as $|0\rangle\langle 0|$ can be turned into a state with off-diagonal elements, which is a consequence of the non-diagonal measurement operator.

Appendix I: Alternative formulation

Up to now we have focused on the dynamics of the energy reservoir, and formulated all our results in terms of channels or CPMs induced on this system. Here we shall take a step back and briefly re-examine the structure of these fluctuation theorems from a global point of view.

1. Global invariance

Let us for a moment forget the division into systems, heat baths, and energy reservoirs, and consider one single system with a global Hamiltonian H, and a unitary evolution V that is energy conserving $[H, V] = 0$, and where this system satisfies a a time-reversal symmetry $\mathcal{T}(H) = H$, $\mathcal{T}(V) = V$. For any pair of global measure-

ment operators Q^{i+} and Q^{f+}, it is the case that

$$\begin{aligned}&\mathrm{Tr}\big(Q^{f+}V\mathcal{J}_{\beta H}(Q^{i+})V^\dagger\big)\\ =&\mathrm{Tr}\big[\mathcal{T}\big(Q^{f+}Ve^{-\beta H/2}Q^{i+}e^{-\beta H/2}V^\dagger\big)\big]\\ =&\mathrm{Tr}\big[Q^{i-}e^{-\beta H/2}VQ^{f-}V^\dagger e^{-\beta H/2}\big]\\ =&\mathrm{Tr}\big(Q^{i-}V\mathcal{J}_{\beta H}(Q^{f-})V^\dagger\big),\end{aligned} \tag{I1}$$

where as usual $Q^{i-} := \mathcal{T}(Q^{i+})$ and $Q^{f-} := \mathcal{T}(Q^{f+})$. In other words, (I1) expresses an invariance of the quantity $\mathrm{Tr}\big(Q^fV\mathcal{J}_{\beta H}(Q^i)V^\dagger\big)$ with respect to the transformation $(Q^i, Q^f) \mapsto (Q^{i'}, Q^{f'}) := \big(\mathcal{T}(Q^f), \mathcal{T}(Q^i)\big)$. All our fluctuation relations can in some sense be regarded as special cases of this global invariance, which here emerges from the combination of time-reversal symmetry and energy conservation.

Time-reversal symmetry alone is not enough to derive this invariance; energy conservation also is needed. However, if it would be the case that $V = e^{-itH/\hbar}$, then it follows that $[H, V] = 0$, and the assumption $\mathcal{T}(H) = H$ would automatically yield $\mathcal{T}(V) = V$. Hence, in this case time-reversal symmetry would be enough.

The relation (I1) can be rewritten as the global fluctuation relation (28) in the main text.

2. Factorization of non-interacting degrees of freedom

The global symmetry in (I1) does not explain how we can express fluctuation theorems in terms of channels or CPMs on the relevant and accessible systems. Throughout this investigation we have repeatedly used the fact that the exponential function factorizes over non-interacting degrees of freedom. To be more precise, suppose that the global system be decomposed into two subsystems 1 and 2 (i.e., $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$) and assume a non-interacting Hamiltonian $H = H_1 \otimes \hat{1}_2 + \hat{1}_1 \otimes H_2$, with the consequence that $\mathcal{J}_{\beta H} = \mathcal{J}_{\beta H_1} \otimes \mathcal{J}_{\beta H_2}$. For product measurement operators $Q = Q_1 \otimes Q_2$ this results in the factorization of the partition map $\mathcal{Z}_{\beta H}(Q_1 \otimes Q_2) = \mathcal{Z}_{\beta H_1}(Q_1)\mathcal{Z}_{\beta H_2}(Q_2)$, and thus also for the Gibbs map $\mathcal{G}_{\beta H}(Q_1 \otimes Q_2) = \mathcal{G}_{\beta H_1}(Q_1) \otimes \mathcal{G}_{\beta H_2}(Q_2)$. In other words, the two states $\mathcal{G}_{\beta H_1}(Q_1)$ and $\mathcal{G}_{\beta H_2}(Q_2)$ can be prepared separately on respective system.

One way in which the factorization could fail would be if the global Hamiltonian is interacting, and this is the topic of section K. In section I 4 we briefly discuss another more exotic possibility, namely if the thermal states would not be characterized by the Gibbs states.

3. Inaccessible degrees of freedom

In the context of statistical mechanics we would typically deal with large numbers of degrees of freedom that we have no access to, e.g., the heat bath. This means that we neither have direct access to prepare arbitrary states on these degrees of freedom, nor to make arbitrary measurements on them. (One can of course imagine some form of partial accessibility, but to keep things simple we here assume 'all or nothing'.) Thus imagine that the total system is divided into two subsystem 1 and 2, where 2 is inaccessible to us, and 1 is completely accessible. We also assume that the global Hamiltonian is non-interacting $H = H_1 \otimes \hat{1}_2 + \hat{1}_1 \otimes H_2$.

Since system 2 is inaccessible to us, all available measurement operators are of the form $Q = Q_1 \otimes \hat{1}_2$, i.e., we can only perform the trivial measurement on system 2. Correspondingly, a trivial preparation would be an equilibrium state $G_\beta(H_2)$ on the unaccessible degrees of freedom, with the philosophy is that nature provides equilibrium states 'for free'. Hence, all possible initial states that we would be able to prepare would be of the form $\rho = \rho_1 \otimes G_\beta(H_2)$. Note that this presupposes the non-interacting global Hamiltonian; without a well defined local Hamiltonian H_2 it would generally not be meaningful to speak of $G_\beta(H_2)$ as being the equilibrium state of 2.

To further highlight how the factorization property enters in the treatment of the unaccessible degrees of freedom, let us take a closer look at the global fluctuation relation in (28). Suppose that the forward process would be characterize by a measurement operator of the allowed form $Q^{f+} = Q_1^{f+} \otimes \hat{1}_2$. Then we know that the initial state of the reverse process would be given by $\mathcal{G}_{\beta H}(Q_1^{f-} \otimes \hat{1}_2)$, where we have assumed $\mathcal{T} = \mathcal{T}_1 \otimes \mathcal{T}_2$. Due to the factorization property we know that $\mathcal{G}_{\beta H}(Q_1^{f-} \otimes \hat{1}_2) = \mathcal{G}_{\beta H_1}(Q_1^{f-}) \otimes \mathcal{G}_{\beta H_2}(\hat{1}_2) = \mathcal{G}_{\beta H_1}(Q_1^{f-}) \otimes G_\beta(H_2)$. In other words, trivial measurements on the unaccessible degrees of freedom gets mapped to trivial preparations on these systems, and vice versa. However, if $\mathcal{G}_{\beta H}(Q_1 \otimes \hat{1}_2)$ would not factorize, this would no longer be true.

This line of reasoning may also fail if $\mathcal{G}_{\beta H_2}(\hat{1}_2)$ would *not* correspond to the equilibrium state on system 2. This would be the case if we would choose β in our fluctuation relations to be different from the actual β' of the heat bath. From a purely mathematical point of view, the fluctuation relations are of course valid for all values of β irrespective of whether they correspond to the actual temperature or not. However, in this case $G_{\beta'}(H_2)$ rather than $G_\beta(H_2)$ would be the true equilibrium state. Hence, it would require an active intervention on system 2 to prepare the state $\mathcal{G}_{\beta H_1}(Q_1^{f-}) \otimes G_\beta(H_2)$.

4. The issue with generalized Gibbs maps

As an illustration of the particular role of the exponential function and the Gibbs distribution, let us imagine that we attempt to use some other form of function, e.g., in the context of generalized notions of equilibrium states. More precisely, let g be any reasonable (possibly complex valued) function, and define the generalized map $\mathcal{J}^g_{\beta H}(Q) := g(\beta H)Qg(\beta H)^\dagger$, as well as the general-

ized Gibbs and partition maps

$$\mathcal{G}^g_{\beta H}(Q) := \frac{\mathcal{J}^g_{\beta H}(Q)}{\mathcal{Z}^g_{\beta H}(Q)}, \quad \mathcal{Z}^g_{\beta H}(Q) := \mathrm{Tr}\mathcal{J}^g_{\beta H}(Q).$$

Here $\mathcal{G}_g(\hat{1}) = g(\beta H)g(\beta H)^\dagger/\mathrm{Tr}\big(g(\beta H)g(\beta H)^\dagger\big)$ would presumably take the role of a generalized form of equilibrium state.

One could imagine to construct fluctuation relations for this generalized setup, and it is indeed straightforward to repeat the derivation of (I1) to obtain

$$\mathrm{Tr}\big(Q^{f+}V\mathcal{J}^g_{\beta H}(Q^{i+})V^\dagger\big) = \mathrm{Tr}\big(Q^{i-}V\mathcal{J}^g_{\beta H}(Q^{f-})V^\dagger\big),$$

At first sight this may seem like an endless source of non-standard fluctuation theorems for hypothetical non-Gibbsian distributions. However, since the factorization property fails in the general case, we can for example not reproduce the reasoning in section I 3.

Appendix J: Pre-correlations

Here we provide some further details on the example of section V C in the main text, where we describe a setup similar to the one for the quantum Crooks relation (2), but where we wish to allow for the possibility that S and E are pre-correlated. The quantum Crooks relation (2) is formulated in terms of channels on the energy reservoir E. However, this implicitly assumes that E initially is uncorrelated with the degrees of freedom it is about to interact with. Hence, in the present case we cannot express fluctuation relations in terms of channels or CPMs on E alone, but an alternative is to formulate a fluctuation relation in terms of channels on the joint system SE.

We let the global Hamiltonian be

$$\begin{aligned} H :=& H^i_{SE}\otimes\hat{1}_B\otimes P^i_C + H^f_{SE}\otimes\hat{1}_B\otimes P^f_C \\ &+ \hat{1}_{SE}\otimes H_B\otimes\hat{1}_C \end{aligned}$$

where H^i_{SE} and H^f_{SE} are the initial and final Hamiltonians of the combined system and energy reservoir (where we can let these be non-interacting if we so wish). Moreover, like in Assumptions 2, P^i_C and P^f_C are the projectors onto $\mathrm{Sp}\{|c_{i+}\rangle, |c_{i-}\rangle\}$ and $\mathrm{Sp}\{|c_{f+}\rangle, |c_{f-}\rangle\}$, respectively.

The total time-reversal is of the form $\mathcal{T} := \mathcal{T}_{SE}\otimes\mathcal{T}_B\otimes\mathcal{T}_C$. (This is different from the decomposition $\mathcal{T}_{SBC}\otimes\mathcal{T}_E$ that we use in Section C.) We assume

$$\mathcal{T}_B(H_B) = H_B,\ \mathcal{T}_{SE}(H^i_{SE}) = H^i_{SE},\ \mathcal{T}_{SE}(H^f_{SE}) = H^f_{SE}.$$

The two latter conditions can for example be achieved if $\mathcal{T}_{SE}$ is the transpose with respect to an orthonormal basis where both H^i_{SE} and H^f_{SE} are represented as real matrices. We furthermore let $\mathcal{T}_C$ be such that

$$\begin{aligned} \mathcal{T}_C(|c_{i+}\rangle\langle c_{i+}|) =& |c_{i-}\rangle\langle c_{i-}|, \\ \mathcal{T}_C(|c_{f+}\rangle\langle c_{f+}|) =& |c_{f-}\rangle\langle c_{f-}|. \end{aligned}$$

We assume a global unitary evolution operator V that is energy conserving $[V, H] = 0$, time-reversal symmetric $\mathcal{T}(V) = V$, and satisfies perfect control

$$V[\hat{1}_{SBE}\otimes|c_{i+}\rangle\langle c_{i+}|] = [\hat{1}_{SBE}\otimes|c_{f+}\rangle\langle c_{f+}|]V.$$

Under these conditions one can show, much along the same lines as our previous derivations, that the channels

$$\begin{aligned} \overline{\mathcal{F}}_+(\chi) :=& \mathrm{Tr}_{CB}(V[|c_{i+}\rangle\langle c_{i+}|\otimes G(H_B)\otimes\chi]V^\dagger), \\ \overline{\mathcal{F}}_-(\chi) :=& \mathrm{Tr}_{CB}(V[|c_{f-}\rangle\langle c_{f-}|\otimes G(H_B)\otimes\chi]V^\dagger), \end{aligned}$$

satisfy the relation $\overline{\mathcal{F}}_+ = \mathcal{J}_{\beta H^f_{SE}}\overline{\mathcal{F}}^{\ominus}_-\mathcal{J}^{-1}_{\beta H^i_{SE}}$.

Appendix K: Approximate fluctuation relations

As discussed in section A 2 we have up to now separated the role of the global Hamiltonian H as characterizing energy, from its role as generator of the time evolution. The latter role we have so far assigned to the unitary operator V, with the restriction that it should be energy conserving $[H, V] = 0$. Here we consider the modifications needed to re-join these two roles in the sense that we let $V = e^{-itH/\hbar}$ (which automatically satisfies the condition for energy conservation $[H, V] = 0$).

Some issues appear when we try to fit $V = e^{-itH/\hbar}$ with our previous assumptions on the structure of H. For example, in Assumptions 2 we explicitly assumed that the global Hamiltonian is of the form $H = H_{S'C}\otimes\hat{1}_E + \hat{1}_{S'C}\otimes H_E$, where $H_{S'C} = H^i_{S'}\otimes P^i_C + H^f_{S'}\otimes P^f_C + H^\perp$. We furthermore assumed that the initial state of the control system in the forward process has support in the subspace onto which P^i_C projects. If the global evolution is given by $V = e^{-itH/\hbar}$, this implies that the the control system will never leave the initial subspace, and thus fail to satisfy the assumption of perfect control. In other words, we cannot obtain the relevant dynamics with these combinations of assumptions.

This particular issue does not apply to the setting of the conditional fluctuation relations in Assumptions 4, since we there neither assume a special structure of $H_{S'C}$, nor of the initial state. However, in the conditional setup we still assume that the global Hamiltonian is to the form $H = H_{\widetilde{S}}\otimes\hat{1}_E + \hat{1}_{\widetilde{S}}\otimes H_E$. In other words, we assume that there is no interaction between $\widetilde{S} = S'C$ and the energy reservoir E. Needless to say, if the global evolution would be given by $V = e^{-itH/\hbar}$, then the state of the energy reservoir would be left unaffected by whatever happens in $S'C$, and the whole notion of the energy reservoir thus becomes meaningless. Hence, we again find that an evolution of the form $V = e^{-itH/\hbar}$ clashes with our general assumptions in the sense that it generates a trivial dynamics.

1. A general notion of approximate fluctuation relations

To highlight the general structure we consider a separation into two anonymous subsystems 1 and 2. The reason for this is that we will consider different ways to partition the subsystems S, B, C, and E.

a. The approximation

In section I 2 we pointed out that for non-interacting Hamiltonians $H = H_1 \otimes \hat{1}_2 + \hat{1}_1 \otimes H_2$, the function $\mathcal{J}_{\beta H}$ satisfies the factorization property $\mathcal{J}_{\beta H}(Q_1 \otimes Q_2) = \mathcal{J}_{\beta H_1}(Q_1) \otimes \mathcal{J}_{\beta H_2}(Q_2)$. Since we in this section abandon these convenient non-interacting Hamiltonians, the question is what is supposed to replace them. Intuitively, the idea is that for positive operators with *suitable* support, the action of the global Hamiltonian H can be approximated by $H_1^i \otimes \hat{1}_2 + \hat{1}_1 \otimes H_2^i$, for some local Hamiltonians H_1^i and H_2^i. Similarly, for another suitable class of operators, the action of the global Hamiltonian can be approximated by $H_1^f \otimes \hat{1}_2 + \hat{1}_1 \otimes H_2^f$, for Hamiltonians H_1^f and H_2^f. (For the sake of generality and flexibility we here allow different initial and final Hamiltonians on both subsystems.)

The more exact formulation is based on the $\mathcal{J}_{\beta H}$ map. We assume a product time reversal $\mathcal{T} = \mathcal{T}_1 \otimes \mathcal{T}_2$ and local approximate Hamiltonians H_1^i, H_1^f, H_2^i, H_2^f such that

$$\begin{aligned} &\mathcal{T}(H) = H, \\ &\mathcal{T}_1(H_1^i) = H_1^i, \quad \mathcal{T}_1(H_1^f) = H_1^f, \\ &\mathcal{T}_2(H_2^i) = H_2^i, \quad \mathcal{T}_2(H_2^f) = H_2^f, \end{aligned} \tag{K1}$$

Suppose that for the measurement operators $Q_1^{i+}, Q_2^{i+}, Q_1^{f+}, Q_2^{f+}$ the approximate factorization holds

$$\begin{aligned} \mathcal{J}_{\beta H}(Q_1^{i+} \otimes Q_2^{i+}) \approx& \mathcal{J}_{\beta H_1^i}(Q_1^{i+}) \otimes \mathcal{J}_{\beta H_2^i}(Q_2^{i+}), \\ \mathcal{J}_{\beta H}(Q_1^{f+} \otimes Q_2^{f+}) \approx& \mathcal{J}_{\beta H_1^f}(Q_1^{f+}) \otimes \mathcal{J}_{\beta H_2^f}(Q_2^{f+}). \end{aligned} \tag{K2}$$

The idea is that under these conditions we should obtain the following approximate fluctuation relation

$$\begin{aligned} &\mathcal{Z}_{\beta H_1^i}(Q_1^i)\mathcal{Z}_{\beta H_2^i}(Q_2^i)P^V_{\beta H^i}[Q_1^{i+} \otimes Q_2^{i+} \to Q_1^{f+} \otimes Q_2^{f+}] \\ \approx& \mathcal{Z}_{\beta H_1^i}(Q_1^f)\mathcal{Z}_{\beta H_2^i}(Q_2^f)P^V_{\beta H^f}[Q_1^{f-} \otimes Q_2^{f-} \to Q_1^{i-} \otimes Q_2^{i-}], \end{aligned} \tag{K3}$$

where $H^i := H_1^i \otimes \hat{1}_2 + \hat{1}_1 \otimes H_2^i$, $H^f := H_1^f \otimes \hat{1}_2 + \hat{1}_1 \otimes H_2^f$, and where $V = e^{-itH/\hbar}$ for some $t \in \mathbb{R}$, and $Q_1^{i-} := \mathcal{T}_1(Q_1^{i+})$, $Q_2^{i-} := \mathcal{T}_2(Q_2^{i+})$, $Q_1^{f-} := \mathcal{T}_1(Q_1^{f+})$, $Q_2^{f-} := \mathcal{T}_2(Q_2^{f+})$.

It is straightforward to make an informal 'derivation' of (K3), which combines the approximations in (K2) with the assumed properties (K1) of the time reversal together with $[V, H] = 0$ and $\mathcal{T}(V) = V$

$$\begin{aligned} &\mathrm{Tr}\big([Q_1^{f+} \otimes Q_2^{f+}]V[\mathcal{J}_{\beta H_1^i}(Q_1^{i+}) \otimes \mathcal{J}_{\beta H_2^i}(Q_2^{i+})]V^\dagger\big) \\ \approx& \mathrm{Tr}\big([Q_1^{f+} \otimes Q_2^{f+}]V\mathcal{J}_{\beta H}(Q_1^{i+} \otimes Q_2^{i+})V^\dagger\big) \\ =& \mathrm{Tr}\big([Q_1^{i-} \otimes Q_2^{i-}]\mathcal{J}_{\beta H}(V[Q_1^{f-} \otimes Q_2^{f-}]V^\dagger)\big) \\ =& \mathrm{Tr}\big([Q_1^{i-} \otimes Q_2^{i-}]V\mathcal{J}_{\beta H}(Q_1^{f-} \otimes Q_2^{f-})V^\dagger\big) \\ \approx& \mathrm{Tr}([Q_1^{i-} \otimes Q_2^{i-}]V[\mathcal{J}_{\beta H_1^f}(Q_1^{f-}) \otimes \mathcal{J}_{\beta H_2^f}(Q_2^{f-})]V^\dagger). \end{aligned}$$

b. Quantitative formulation of the approximation

Here we consider one way to make the approximation as expressed by (K2) and (K3) quantitative. We here assume that the underlying Hilbert spaces are finite-dimensional. Define

$$\begin{aligned} &d^H_{H_1,H_2}(Q_1, Q_2) \\ &:= \big\|\mathcal{J}_{\beta H_1}(Q_1) \otimes \mathcal{J}_{\beta H_2}(Q_2) - \mathcal{J}_{\beta H}(Q_1 \otimes Q_2)\big\|_1, \end{aligned} \tag{K4}$$

where we use the standard operator norm $\|Q\| := \sup_{\|\psi\|=1} \|Q|\psi\rangle\|$ and the trace norm $\|Q\|_1 := \mathrm{Tr}\sqrt{Q^\dagger Q}$.

By Lemma 6, we can conclude that $\|Q_1^{i-}\| = \|Q_1^{i+}\|$ and analogous for $Q_2^{i\pm}$, $Q_1^{f\pm}$, $Q_2^{f\pm}$. Moreover, one can confirm that $d^H_{H_1^i,H_2^i}(Q_1^{i-}, Q_2^{i-}) = d^H_{H_1^i,H_2^i}(Q_1^{i+}, Q_2^{i+})$ and $d^H_{H_1^f,H_2^f}(Q_1^{f-}, Q_2^{f-}) = d^H_{H_1^f,H_2^f}(Q_1^{f+}, Q_2^{f+})$.

As a bit of a technical side-remark one may note that none of the proofs explicitly use the assumption that $V = e^{-itH/\hbar}$. We do still only rely on $[V, H] = 0$ and $\mathcal{T}(V) = V$. However, due to the more forgiving structure we can now assume that $V = e^{-itH/\hbar}$, and yet have a non-trivial evolution on the energy reservoir.

Proposition 10. *Let H, H_1^i, H_2^i, H_1^f, H_2^f be Hermitian operators, let $\mathcal{T}, \mathcal{T}_1, \mathcal{T}_2$ be time reversals that satisfy the conditions in (K1), and let V be a unitary operator such that $[H, V] = 0$ and $\mathcal{T}(V) = V$ (which in particular allows us to choose $V = e^{-itH/\hbar}$ for some $t \in \mathbb{R}$). Then*

$$\begin{aligned} &\Big|\mathcal{Z}_{\beta H_1^i}(Q_1^i)\mathcal{Z}_{\beta H_2^i}(Q_2^i)P^V_{\beta H^i}[Q_1^{i+} \otimes Q_2^{i+} \to Q_1^{f+} \otimes Q_2^{f+}] \\ &- \mathcal{Z}_{\beta H_1^i}(Q_1^f)\mathcal{Z}_{\beta H_2^i}(Q_2^f)P^V_{\beta H^f}[Q_1^{f-} \otimes Q_2^{f-} \to Q_1^{i-} \otimes Q_2^{i-}]\Big| \\ &\le \|Q_1^{f+}\|\|Q_2^{f+}\|d^H_{H_1^i,H_2^i}(Q_1^{i+}, Q_2^{i+}) \\ &+ \|Q_1^{i+}\|\|Q_2^{i+}\|d^H_{H_1^f,H_2^f}(Q_1^{f+}, Q_2^{f+}), \end{aligned}$$

where $H^i := H_1^i \otimes \hat{1}_2 + \hat{1}_1 \otimes H_2^i$, $H^f := H_1^f \otimes \hat{1}_2 + \hat{1}_1 \otimes H_2^f$, and where $Q_1^{i-} := \mathcal{T}_1(Q_1^{i+})$, $Q_2^{i-} := \mathcal{T}_2(Q_2^{i+})$, $Q_1^{f-} := \mathcal{T}_1(Q_1^{f+})$, $Q_2^{f-} := \mathcal{T}_2(Q_2^{f+})$.

Proof. As the first step one can confirm the identity

$$\begin{aligned} &\mathrm{Tr}\big([Q_1^{f+} \otimes Q_2^{f+}]V\mathcal{J}_{\beta H}(Q_1^{i+} \otimes Q_2^{i+})V^\dagger\big) \\ =& \mathrm{Tr}\big([Q_1^{i-} \otimes Q_2^{i-}]V\mathcal{J}_{\beta H}(Q_1^{f-} \otimes Q_2^{f-})V^\dagger\big). \end{aligned} \tag{K5}$$

For a more compact notation let $\sigma_i := \mathcal{J}_{\beta H_1^i}(Q_1^{i+}) \otimes \mathcal{J}_{\beta H_2^i}(Q_2^{i+})$ and $\sigma_f := \mathcal{J}_{\beta H_1^f}(Q_1^{f-}) \otimes \mathcal{J}_{\beta H_2^f}(Q_2^{f-})$. Then

$$\begin{aligned}
&\left|\mathrm{Tr}([Q_1^{f+}\otimes Q_2^{f+}]V\sigma_i V^\dagger) - \mathrm{Tr}([Q_1^{i-}\otimes Q_2^{i-}]V\sigma_f V^\dagger)\right| \\
&\qquad \text{[By (K5) and the triangle inequality]} \\
&\leq \left|\mathrm{Tr}\Big(V^\dagger[Q_1^{f+}\otimes Q_2^{f+}]V\big(\sigma_i - \mathcal{J}_{\beta H}(Q_1^{i+}\otimes Q_2^{i+})\big)\Big)\right| \\
&+ \left|\mathrm{Tr}\Big(V^\dagger[Q_1^{i-}\otimes Q_2^{i-}]V\big(\mathcal{J}_{\beta H}(Q_1^{f-}\otimes Q_2^{f-}) - \sigma_f\big)\Big)\right| \\
&\qquad \text{[By the general relation } |\mathrm{Tr}(AQ)| \leq \|A\|\|Q\|_1] \\
&\leq \|Q_1^{f+}\|\|Q_2^{f+}\| d^H_{H_1^i,H_2^i}(Q_1^{i+},Q_2^{i+}) \\
&+ \|Q_1^{i+}\|\|Q_2^{i+}\| d^H_{H_1^f,H_2^f}(Q_1^{f+},Q_2^{f+}).
\end{aligned}$$

□

2. Approximate conditional fluctuation relations

Here we consider an approximate versions of the conditional fluctuation relations in section H. The general case is treated in the following subsection, and in the next we consider a more concrete special case of a particle as a control system.

a. *The general case*

Here we identify system 1 of the previous section with $\widetilde{S}$, and system 2 with E. We assume

$$\begin{aligned}
&\mathcal{T}(H) = H, \\
&\mathcal{T}_{\widetilde{S}}(H^i_{\widetilde{S}}) = H^i_{\widetilde{S}}, \quad \mathcal{T}_{\widetilde{S}}(H^f_{\widetilde{S}}) = H^f_{\widetilde{S}}, \\
&\mathcal{T}_E(H^i_E) = H^i_E, \quad \mathcal{T}_E(H^f_E) = H^f_E.
\end{aligned} \tag{K6}$$

Similar to the conditional fluctuation relations in section H we define CPMs on the energy reservoir conditioned on the successful control measurements

$$\begin{aligned}
\tilde{\mathcal{F}}_+(\sigma) :=& \mathrm{Tr}_{\widetilde{S}}([Q^{f+}_{\widetilde{S}}\otimes \hat{1}_E]V[\mathcal{G}_{\beta H^i_{\widetilde{S}}}(Q^{i+}_{\widetilde{S}})\otimes\sigma]V^\dagger), \\
\tilde{\mathcal{F}}_-(\sigma) :=& \mathrm{Tr}_{\widetilde{S}}([Q^{i-}_{\widetilde{S}}\otimes \hat{1}_E]V[\mathcal{G}_{\beta H^f_{\widetilde{S}}}(Q^{f-}_{\widetilde{S}})\otimes\sigma]V^\dagger),
\end{aligned} \tag{K7}$$

One can verify that

$$\begin{aligned}
P^{\tilde{\mathcal{F}}_+}_{\beta H^i_E}[Q^{i+}_E \to Q^{f+}_E] &= P^V_{\beta H^i}[Q^{i+}_{\widetilde{S}}\otimes Q^{i+}_E \to Q^{f+}_{\widetilde{S}}\otimes Q^{f+}_E], \\
P^{\tilde{\mathcal{F}}_-}_{\beta H^f_E}[Q^{f-}_E \to Q^{i-}_E] &= P^V_{\beta H^f}[Q^{f-}_{\widetilde{S}}\otimes Q^{f-}_E \to Q^{i-}_{\widetilde{S}}\otimes Q^{i-}_E],
\end{aligned}$$

which together with Proposition 10 yield

$$\begin{aligned}
&\Big|\mathcal{Z}_{\beta H^i_{\widetilde{S}}}(Q^i_{\widetilde{S}})\mathcal{Z}_{\beta H^i_E}(Q^i_E)P^{\tilde{\mathcal{F}}_+}_{\beta H^i_E}[Q^{i+}_E \to Q^{f+}_E] \\
&- \mathcal{Z}_{\beta H^f_{\widetilde{S}}}(Q^f_{\widetilde{S}})\mathcal{Z}_{\beta H^i_E}(Q^f_E)P^{\tilde{\mathcal{F}}_-}_{\beta H^f_E}[Q^{f-}_E \to Q^{i-}_E]\Big| \\
&\leq \|Q^{f+}_{\widetilde{S}}\|\|Q^{f+}_E\| d^H_{H^i_{\widetilde{S}},H^i_E}(Q^{i+}_{\widetilde{S}},Q^{i+}_E) \\
&+ \|Q^{i+}_{\widetilde{S}}\|\|Q^{i+}_E\| d^H_{H^f_{\widetilde{S}},H^f_E}(Q^{f+}_{\widetilde{S}},Q^{f+}_E).
\end{aligned} \tag{K8}$$

With (K8) as starting point one can also obtain an approximate fluctuation relation in terms of channels, i.e., more in the spirit of (25). Let us define the following measure of difference between maps

$$\mathrm{Diff}(\phi_1,\phi_2) := \sup_{Q^i\geq 0, Q^f\geq 0} \frac{\left|\mathrm{Tr}\big(Q^f\phi_2(Q^i)\big) - \mathrm{Tr}\big(Q^f\phi_1(Q^i)\big)\right|}{\|Q^i\|\|Q^f\|}. \tag{K9}$$

This particular choice of measure has no deeper reason than that it makes the derivation simple.

$$\begin{aligned}
D_i(Q^{i+}_{\widetilde{S}}) &:= \sup_{Q\geq 0}\frac{1}{\|Q\|} d^H_{H^i_{\widetilde{S}},H^i_E}(Q^{i+}_{\widetilde{S}},Q), \\
D_f(Q^{f+}_{\widetilde{S}}) &:= \sup_{Q\geq 0}\frac{1}{\|Q\|} d^H_{H^f_{\widetilde{S}},H^f_E}(Q^{f+}_{\widetilde{S}},Q).
\end{aligned} \tag{K10}$$

By the approximate conditional fluctuation relation (K8) it follows that

$$\begin{aligned}
&\mathrm{Diff}\big(\mathcal{Z}_{\beta H^i_{\widetilde{S}}}(Q^i_{\widetilde{S}})\tilde{\mathcal{F}}_+\circ\mathcal{J}_{\beta H^i_E}\,,\ \mathcal{Z}_{\beta H^f_{\widetilde{S}}}(Q^f_{\widetilde{S}})\mathcal{J}_{\beta H^f_E}\circ\tilde{\mathcal{F}}^{\ominus}_-\big) \\
&\qquad \leq \|Q^{f+}_{\widetilde{S}}\| D_i(Q^{i+}_{\widetilde{S}}) + \|Q^{i+}_{\widetilde{S}}\| D_f(Q^{f+}_{\widetilde{S}}).
\end{aligned} \tag{K11}$$

Hence, this is a quantitative version of the approximate conditional relation (34) in the main text.

In (K8) we only consider the error for a specific pair Q^{i+}_E, Q^{f+}_E, while in (K11) we ask for the worst case error over the entire set of positive semi definite operators. It may very well be the case that pointwise error can be small for some specific choice of operators, while uniform error would be large. Hence, the formulation via the transition probabilities can be more 'forgiving' than the formulation via channels. This should be compared with the non-approximate case, where the choice largely is a matter of convenience.

b. *The special case of a control-particle*

Imagine a joint Hamiltonian of the form

$$H = \frac{1}{2M_C}\hat{P}_C^2\otimes\hat{1}_{S'E} + H_{S'E}(\hat{X}_C). \tag{K12}$$

Here M_C is the mass and $\hat{X}_C, \hat{P}_C$ are the canonical position and momentum operators of the control particle. Assume furthermore that $H_{S'E}(x)$ is of the form

$$H_{S'E}(x) = \begin{cases} H^i_{S'}\otimes\hat{1}_E + \hat{1}_{S'}\otimes H^i_E, & x\leq x_i, \\ H^f_{S'}\otimes\hat{1}_E + \hat{1}_{S'}\otimes H^f_E, & x\geq x_f, \end{cases} \tag{K13}$$

while in the interval $[x_i, x_f]$ there is some non-trivial dependence on x (where we assume $x_i < x_f$). The Hamiltonians $H^i_{\widetilde{S}}$ and $H^f_{\widetilde{S}}$ would in this case be

$$\begin{aligned}
H^i_{\widetilde{S}} :=& \frac{1}{2M_C}\hat{P}_C^2\otimes\hat{1}_{S'} + \hat{1}_C\otimes H^i_{S'}, \\
H^f_{\widetilde{S}} :=& \frac{1}{2M_C}\hat{P}_C^2\otimes\hat{1}_{S'} + \hat{1}_C\otimes H^f_{S'}.
\end{aligned} \tag{K14}$$

Hence, outside the 'interaction region' $[x_i, x_f]$ all the three systems systems S', E, and C are non-interacting.

For the sake of illustration we consider the special case of measurement operators of the form

$$Q^{i\pm}_{\widetilde{S}} := Q^{i\pm}_C \otimes Q^{i\pm}_{S'}, \quad Q^{f\pm}_{\widetilde{S}} := Q^{f\pm}_C \otimes Q^{f\pm}_{S'}, \tag{K15}$$

where $Q^{i\pm}_{S'}$, $Q^{f\pm}_{S'}$ are measurement operators on S', and where $Q^{i\pm}_C$, $Q^{f\pm}_C$ are measurement operators on C which are 'concentrated' in the regions $(-\infty, x_i]$ and $[x_f, +\infty)$, respectively. It seems intuitively reasonable that the further down from x_i that Q^{i+}_C is supported, the better is the approximation

$$\begin{aligned}
&\mathcal{J}_{\beta H}(Q^{i+}_{\widetilde{S}} \otimes Q) \\
&\approx \mathcal{J}_{\beta H^i_{\widetilde{S}}}(Q^{i+}_{\widetilde{S}}) \otimes \mathcal{J}_{\beta H^i_E}(Q) \\
&= \mathcal{J}_{\beta \hat{K}}(Q^{i+}_C) \otimes \mathcal{J}_{\beta H^i_{S'}}(Q^{i+}_{S'}) \otimes \mathcal{J}_{\beta H^i_E}(Q),
\end{aligned} \tag{K16}$$

where we have introduced the notation $\hat{K} := \hat{P}^2_C/(2M_C)$, and used the fact that (K14) defines non-interacting Hamiltonians between S' and C.

The approximate fluctuation relation (K11) takes the form

$$\begin{aligned}
&\mathrm{Diff}\Big(\mathcal{Z}_{\beta\hat{K}}(Q^{i+}_C)\mathcal{Z}_{H^i_{S'}}(Q^{i+}_{S'})\tilde{\mathcal{F}}_+ \circ \mathcal{J}_{\beta H^i_E}, \\
&\qquad \mathcal{Z}_{\beta\hat{K}}(Q^{f+}_C)\mathcal{Z}_{H^f_{S'}}(Q^{f+}_{S'})\mathcal{J}_{\beta H^f_E} \circ \tilde{\mathcal{F}}^{\ominus}_-\Big) \\
&\leq \|Q^{f+}_C\|\|Q^{f+}_{S'}\| D_i(Q^{i+}_C \otimes Q^{i+}_{S'}) \\
&\qquad + \|Q^{i+}_C\|\|Q^{i+}_{S'}\| D_f(Q^{f+}_C \otimes Q^{f+}_{S'}).
\end{aligned}$$

Note that if Q^{f+}_C is a spatial translation of Q^{i+}_C, i.e., such that $Q^{f+}_C = e^{ir\hat{P}_C} Q^{i+}_C e^{-ir\hat{P}_C}$ for some $r \in \mathbb{R}$, then $\mathcal{Z}_{\beta\hat{K}}(Q^{i+}_C) = \mathcal{Z}_{\beta\hat{K}}(Q^{f+}_C)$.

One may wonder how to choose the time-reversals $\mathcal{T}_{\widetilde{S}}$ and $\mathcal{T}_E$. The simplest case is if there exist orthonormal complete bases of $\mathcal{H}_{S'}$ and $\mathcal{H}_E$, such that the family of Hamiltonians $H_{S'E}(x)$, $H^i_{S'}$, $H^f_{S'}$, H^i_E, and H^f_E are real valued matrices in these bases. In this case we can choose $\mathcal{T}_E$ and $\mathcal{T}_{S'}$ as transpositions with respect to these bases, and $\mathcal{T}_C$ as the transposition with respect to the position representation, and let $\mathcal{T}_{\widetilde{S}} = \mathcal{T}_C \otimes \mathcal{T}_{S'}$. This guarantees that (K6) holds.

As a further comment one may observe that what we assign as being the energy reservoir is largely a matter of choice. In this particular example it is clear that also the control particle can donate energy.

3. Joint control and energy reservoir

Here we turn to a setting where the control and the energy reservoir are one and the same system. Hence, instead of dividing the global system between E and $S'C$, we here divide it into CE and S'.

a. *The general case*

Here we let system 1 (in section K 1 a) be S' and system 2 be CE. We assume

$$\begin{aligned}
&\mathcal{T}(H) = H, \\
&\mathcal{T}_{S'}(H^i_{S'}) = H^i_{S'}, \quad \mathcal{T}_{S'}(H^f_{S'}) = H^f_{S'}, \\
&\mathcal{T}_{CE}(H^i_{CE}) = H^i_{CE}, \quad \mathcal{T}_{CE}(H^f_{CE}) = H^f_{CE},
\end{aligned} \tag{K17}$$

as well as $[V, H] = 0$ and $\mathcal{T}(V) = V$. By Proposition 10

$$\begin{aligned}
&\Big|\mathcal{Z}_{\beta H^i_{S'}}(Q^i_{S'})\mathcal{Z}_{\beta H^i_{CE}}(Q^i_{CE})\mathcal{P}^+ \\
&- \mathcal{Z}_{\beta H^f_{S'}}(Q^f_{S'})\mathcal{Z}_{\beta H^f_{CE}}(Q^f_{CE})\mathcal{P}^-\Big| \\
&\leq \|Q^{f+}_{S'}\|\|Q^{f+}_{CE}\| d^H_{H^i_{S'}, H^i_{CE}}(Q^{i+}_{S'}, Q^{i+}_{CE}) \\
&+ \|Q^{i+}_{S'}\|\|Q^{i+}_{CE}\| d^H_{H^f_{S'}, H^f_{CE}}(Q^{f+}_{S'}, Q^{f+}_{CE}),
\end{aligned} \tag{K18}$$

where we introduce the following short hand notation for the transition probabilities

$$\begin{aligned}
\mathcal{P}^+ :=& P^V_{\beta H^i}[Q^{i+}_{S'} \otimes Q^{i+}_{CE} \to Q^{f+}_{S'} \otimes Q^{f+}_{CE}], \\
\mathcal{P}^- :=& P^V_{\beta H^f}[Q^{f-}_{S'} \otimes Q^{f-}_{CE} \to Q^{i-}_{S'} \otimes Q^{i-}_{CE}].
\end{aligned} \tag{K19}$$

Hence, (K18) is the quantitative version of (37) in the main text.

b. *A single particle as both control and energy reservoir*

In section K 2 b we considered the case of a particle whose motion implements the time-dependent Hamiltonian in Crooks relation. There we regarded the degrees of freedom of the particle as separate from the designated energy reservoir. This made it possible to express CPMs on the energy reservoir conditioned on the control measurements on the control particle. An intuitively reasonable alternative would be that the motion of the control particle also fuels the process, i.e., it is the initial kinetic energy of the particle that that drives the whole non-equilibrium process. The global Hamiltonian is

$$H = \frac{1}{2M_{CE}}\hat{P}^2_{CE} \otimes \hat{1}_{S'} + H_{S'}(\hat{X}_{CE}), \tag{K20}$$

where M_{CE} is the mass, and $\hat{X}_{CE}, \hat{P}_{CE}$ are the canonical position and momentum operators of the particle.

We assume

$$H_{S'}(x) = \begin{cases} H^i_{S'}, & x \leq x_i, \\ H^f_{S'}, & x \geq x_f, \end{cases} \tag{K21}$$

i.e., the Hamiltonian for S' is constant outside the interaction region. Since the particle moreover is free outside the interaction region, we can take the initial and final approximate Hamiltonians for CE

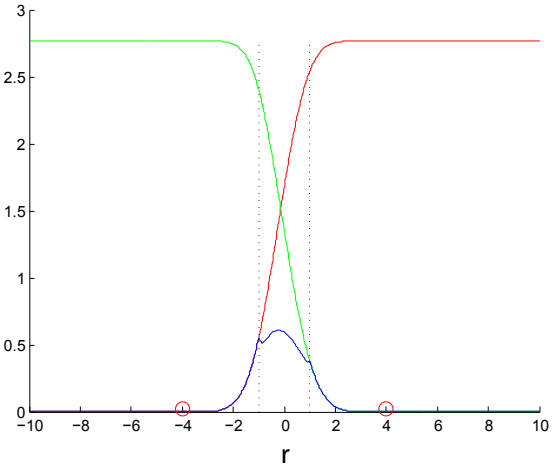

FIG. 10: **Approximate factorization.** To assess the quality of the approximate factorization we evaluate $d^H_{\hat{K},E_0\sigma_z/2}(\hat{1}_{S'}\otimes|\alpha\rangle\langle\alpha|)$ (red curve), $d^H_{\hat{K},E_0\sigma_x/4}(\hat{1}_{S'}\otimes|\alpha\rangle\langle\alpha|)$ (green curve), $d^H_{\hat{K},E_0\overline{n}(ry_0)\cdot\overline{\sigma}/2}(\hat{1}_{S'}\otimes|\alpha\rangle\langle\alpha|)$ (blue curve) for $\alpha := r+2i$ with $r\in[-10,10]$. The dotted black lines correspond to the borders of the interaction region. The red circles are the positions of the coherent states that gives the measurement operators $Q^{i+}_{CE} = |-4+2i\rangle\langle -4+2i|$, $Q^{f+}_{CE} = |4+2i\rangle\langle 4+2i|$.

to be $H^i_{CE} = H^f_{CE} = \frac{1}{2M_{CE}}\hat{P}^2_{CE} =: \hat{K}$. For operators Q^{i+}_{CE} and Q^{f+}_{CE} that are well localized outside the interaction region it seems reasonable that $\mathcal{J}_{\beta H}(Q^{i+}_{S'}\otimes Q^{i+}_{CE}) \approx \mathcal{J}_{\beta H^i_{S'}}(Q^{i+}_{S'})\otimes\mathcal{J}_{\beta\hat{K}}(Q^{i+}_{CE})$ and $\mathcal{J}_{\beta H}(Q^{f+}_{S'}\otimes Q^{f+}_{CE}) \approx \mathcal{J}_{\beta H^f_{S'}}(Q^{f+}_{S'})\otimes\mathcal{J}_{\beta\hat{K}}(Q^{f+}_{CE})$. Under these conditions we thus get the approximate fluctuation relation $\mathcal{Z}_{\beta H^i_{S'}}(Q^{i+}_{S'})\mathcal{Z}_{\beta\hat{K}}(Q^{i+}_{CE})\mathcal{P}^+ \approx \mathcal{Z}_{\beta H^f_{S'}}(Q^{f-}_{S'})\mathcal{Z}_{\beta\hat{K}}(Q^{f-}_{CE})\mathcal{P}^-$, with the quantitative version in (K18).

c. *Numerical evaluation*

To make the approximate fluctuation relations a bit more concrete we here make a numerical evaluation of a special case of the combined control energy reservoir particle in the previous section. We consider a single particle of mass M that is restricted to move along the y-axis, and this spatial degree of freedom is taken as the combined control-energy reservoir CE. The particle also carries a magnetic moment corresponding to a spin-half degree of freedom, which we interpret as system S'. (A single spin is of course somewhat ridiculous regarded as a combined system and heat bath, but this example only serves to illustrate the formalism, for which the sizes of the participating systems do not matter.) We assume that the spin interacts with an external magnetic field that is time-independent, but is a function of y. The total Hamiltonian can be expressed in terms of the differential operator

$$H = -\frac{\hbar^2}{2M}\frac{d^2}{dy^2} + \frac{1}{2}E_0\overline{\sigma}_{S'}\cdot\overline{n}(y), \qquad \text{(K22)}$$

where $\overline{n}(y)$ determines the strength and direction of the external magnetic field as a function of y, and where $\overline{\sigma}_{S'} = (\sigma^{(x)}_{S'},\sigma^{(y)}_{S'},\sigma^{(z)}_{S'})$ are the Pauli spin operators, with $\sigma^{(x)} = |0\rangle\langle 1| + |1\rangle\langle 0|$, $\sigma^{(y)} = i|0\rangle\langle 1| - i|1\rangle\langle 0|$, $\sigma^{(z)} = |1\rangle\langle 1| - |0\rangle\langle 0|$, with $\{|0\rangle,|1\rangle\}$ being the eigenbasis of $\sigma^{(z)}$. For $\|\overline{n}\| = 1$ it follows that E_0 is the excitation energy of the spin.

To get a particularly simple model we here assume that $\overline{n}(y) = (0,0,1)$ for $y < -y_0$, $\overline{n}(y) = [3/4 - y/(4y_0)]\big(\sin(\frac{\pi(y+y_0)}{4y_0}),0,\cos(\frac{\pi(y+y_0)}{4y_0})\big)$ for $-y_0\le y\le y_0$, and $\overline{n}(y) = (1/2,0,0)$ for $y_0 < y$. Hence, for all positions below $-y_0$ the magnetic field is directed along the z-axis. Within the interaction region $[-y_0,y_0]$ the field rotates in the xz-plane until it aligns with the x-axis at y_0, simultaneously as it decreases in strength to the half. We choose $H^i_{CE} = H^f_{CE} = \hat{P}^2/(2M) =: \hat{K}$, $H^i_{S'} := E_0\sigma_z/2$, and $H^f_{S'} := E_0\sigma_x/4$.

Define $\mathcal{T}_{CE}$ to be the transpose with respect to the coordinate representation, and thus $\mathcal{T}_{CE}(\hat{K}) = \hat{K}$. For the chosen $\overline{n}(y)$, the Hamiltonian $\overline{n}(y)\cdot\overline{\sigma}$ is represented as a real valued matrix in the eigenbasis of σ_z. Hence, $\mathcal{T}_{S'}$ can be chosen as the transpose with respect to the eigenbasis of σ_z, and thus $\mathcal{T}_{S'}\big(\overline{n}(y)\cdot\overline{\sigma}\big) = \overline{n}(y)\cdot\overline{\sigma}$, $\mathcal{T}_{S'}(H^i_{S'}) = H^i_{S'}$, and $\mathcal{T}_{S'}(H^f_{S'}) = H^f_{S'}$. We also get $\mathcal{T}(H) = H$.

As measurement operators on CE we choose projectors onto coherent states $Q^{i+}_{CE} := |\alpha_i\rangle\langle\alpha_i|$, $Q^{f+}_{CE} := |\alpha_f\rangle\langle\alpha_f|$, where the corresponding wave-functions are

$$\psi_\alpha(y) = \frac{1}{(2\pi)^{1/4}}\frac{1}{\sqrt{\sigma}}e^{-\mathrm{Im}(\alpha)^2}\exp[-\frac{1}{4}(\frac{y}{\sigma}-2\alpha)^2].$$

Here, σ is the standard deviation, and $2\sigma\mathrm{Re}(\alpha)$ the expectation value, of the corresponding Gaussian distribution $|\psi_\alpha(y)|^2$. (With the coherent state defined as the displaced ground state of a harmonic oscillator, σ is determined by the parameters of the chosen oscillator.) Similarly, $\hbar\mathrm{Im}(\alpha)/\sigma$ is the average momentum of the coherent state. One can confirm that $\mathcal{T}_{CE}(|\alpha\rangle\langle\alpha|) = |\alpha^*\rangle\langle\alpha^*|$, and thus the time-reversal changes the sign of the momentum, but leaves the position intact. For the spin degree of freedom we let the measurement operators be $Q^{i+}_{S'} = Q^{f+}_{S'} = \hat{1}_{S'}$. We choose (somewhat arbitrarily) the parameters such that $\hbar^2/(ME_0y_0^2) = 0.1$, $\beta E_0 = 1$ and $\sigma = y_0/2$. This means that the typical thermal energy kT is equal to the excitation energy of the spin (for $y\le -y_0$), and the width of the wave-packet is of the same order as the size of the interaction region. Figure 10 displays the numerical evaluation of the factorization error $d^H_{H_1,H_2}(\hat{1}_{S'},|\alpha\rangle\langle\alpha|)$ as defined in (K4). Here we choose $H_1 := \hat{K}$, and as H_2 we choose $E_0\sigma_z/2$ (red curve), and $E_0\sigma_x/4$ (green curve), for $\alpha := r+2i$ with $r\in[-10,10]$. Since $\sigma = y_0/2$ it follows that the spatial wave-packet ψ_α

is centered at ry_0. For the sake of comparison we also include the 'local' approximation $H_2 := E_0\overline{n}(ry_0)\cdot\overline{\sigma}/2$ (blue curve), where for a state centered at the location ry_0 we approximate the total Hamiltonian H with the non-interacting Hamiltonian $\hat{K} + E_0\overline{n}(ry_0)\cdot\overline{\sigma}/2$. By construction, the blue curve coincides with the green for $r \leq -1$, and with the red for $1 \leq r$. Maybe unsurprisingly, the local approximation is better than the others inside the interaction region.

We now turn to the test of the approximate fluctuation relation, and for this purpose we choose an evolution time t such that $tE_0/\hbar = 21.5$. (This particular choice happens to make the transition probability for the forward process fairly large.) In figure 11 the final states of the particle in the forward and reverse processes are depicted, where we use the measurement operators $Q^{i+}_{CE} := |-4+2i\rangle\langle -4+2i|$, $Q^{f+}_{CE} := |4+2i\rangle\langle 4+2i|$, and $Q^{i+}_{S'} = Q^{f+}_{S'} := \hat{1}_{S'}$. The transition probabilities [in equation (K19)] of the forward and reverse processes are $\mathcal{P}^+ \approx 0.36$ and $\mathcal{P}^- \approx 0.39$, respectively. The error in the approximate fluctuation relation, as defined by the left hand side of equation (K18) becomes $|Z_i\mathcal{P}^+ - Z_f\mathcal{P}^-| \approx 1.6*10^{-8}$ where $Z_i := Z_\beta(E_0\sigma_z/2)\mathcal{Z}_{\beta\hat{K}}(Q^{i+}_{CE})$ and $Z_f := Z_\beta(E_0\sigma_x/4)\mathcal{Z}_{\beta\hat{K}}(Q^{f-}_{CE})$. An estimate of the relative error is $|Z_i\mathcal{P}^+ - Z_f\mathcal{P}^-|/(|Z_i\mathcal{P}^+| + |Z_f\mathcal{P}^-|) \approx 2.2*10^{-8}$. One can also calculate the upper bound in the right hand side of (K18), which becomes $d^H_{H^i_{S'},H^i_{CE}}(Q^{i+}_{S'}, Q^{i+}_{CE}) + d^H_{H^f_{S'},H^f_{CE}}(Q^{f+}_{S'}, Q^{f+}_{CE}) \approx 1.2*10^{-5}$ (where we use the fact that $\|Q^{i+}_{S'}\| = 1$, $\|Q^{f+}_{S'}\| = 1$, $\|Q^{i+}_{CE}\| = 1$, $\|Q^{f+}_{CE}\| = 1$).

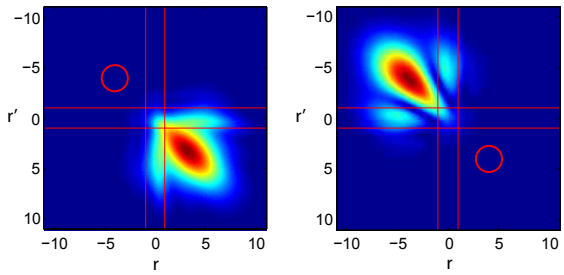

FIG. 11: **The forward and reverse process.** The density operators of the combined control and energy reservoir particle in the position representation (the absolute values of the matrix elements) at the end of the forward (left) and reverse (right) process. The pairs of horizontal and vertical lines shows the borders of the interaction region. The red circles indicate the positions of the initial states $\mathcal{G}_{\beta\hat{K}}(Q^{i+}_{CE})$ and $\mathcal{G}_{\beta\hat{K}}(Q^{f+}_{CE})$ for the initial and final measurement operators $Q^{i+}_{CE} = |-4+2i\rangle\langle -4+2i|$, $Q^{f+}_{CE} = |4+2i\rangle\langle 4+2i|$.
The error in the approximate fluctuation relation is small (approximately $1.6*10^{-8}$), which corresponds to the fact that the measurement operators that are well separated from the interaction region. One should compare this with the final wave packets, where one clearly can see that these have significant weights within the interaction region. This illustrates the fact that it is the properties of the measurement operators, rather than the final states, that matter for the quality of the approximation.
One can also note that the final states are not mirror images of each other. Hence, the symmetry discussed in section I1 does not imply that the wave packets of the forward and reverse process have to be symmetric images of each other.

Appendix L: Some additional remarks

1. Detailed balance

One way of obtaining fluctuation relations in the classical case is via stochastic dynamics that satisfies detailed balance [3, 7, 143]. The reader may have noted that we in our derivations never refer to detailed balance. The reason is that energy conservation and time reversal symmetry in some sense supersedes it, which we give a brief demonstration of here.

Let H_1 and H_2 be non-degenerate Hermitian operators on a finite-dimensional Hilbert space. We let the global Hamiltonian be non-interacting $H = H_1 \otimes \hat{1}_2 + \hat{1}_1 \otimes H_2$, and assume an energy conserving unitary evolution $[H, V] = 0$, and a product time-reversal $\mathcal{T} = \mathcal{T}_1 \otimes \mathcal{T}_2$ with $\mathcal{T}_1(H_1) = H_1$ and $\mathcal{T}_2(H_2) = H_2$. Assuming system 2 is in equilibrium we can define the transition probability of changing the state of system 1 from eigenstate n to n' as

$$p(n'|n) := \mathrm{Tr}([|n'\rangle\langle n'| \otimes \hat{1}_2]V[|n\rangle\langle n| \otimes G(H_2)]V^\dagger). \quad \text{(L1)}$$

By using the energy conservation (and the observation $G(H)[|n\rangle\langle n| \otimes \hat{1}_2] = G_n(H_1)|n\rangle\langle n| \otimes G(H_2)$) it follows that

$$\begin{aligned} p(n'|n)G_n(H_1) =& G_{n'}(H_1)\mathrm{Tr}([|n\rangle\langle n| \otimes \hat{1}_2] \\ & V^\dagger[|n'\rangle\langle n'| \otimes G(H_2)]V). \end{aligned}$$

This is almost what we want, apart from the fact that the evolution is reversed. This is the point where we can make use of the time reversal symmetry (and Lemma 7) to obtain $p(n'|n)G_n(H_1) = p(n|n')G_{n'}(H_1)$, i.e., the transition probability $p(n'|n)$ satisfies detailed balance.

2. Heat baths in the Gibbs state

We do in this investigation often assume that the heat bath initially is in the Gibbs state corresponding to a given temperature. Although not unusual assumption, e.g., in derivations of master equations [144] and [94] (in particular sections 3.6.2.1 and 4.2.2), it is nevertheless worth considering the justification, especially since one may argue that it is not the heat bath *per se* that is Gibbs distributed, but rather systems that are weakly coupled to it. One possible argument would be that the environment can be separated into a 'near environment' that is relevant on the time scale of the experiment, and

a 'far environment' (or 'super bath') that puts the near environment in the Gibbs state (see e.g. the discussions in [10]). Another approach would be to assume that an ideal heat bath in some sense behaves *as if* it is Gibbs distributed. These issues approach the question of thermalization in closed systems, and along these lines one may speculate that typicality [115, 116] could be employed to obtain a more refined analysis of fluctuation relations.

---

[1] C. Bustamante, J. Liphardt, F. Ritort, *The nonequilibrium thermodynamics of small systems*, Phys. Today **58**, 43 (2005).

[2] L. Rondoni, C. Mejía-Monasterio, *Fluctuations in nonequilibrium statistical mechanics: models, mathematical theory, physical mechanisms*, Nonlinearity **20**, R1 (2007).

[3] C. Jarzynski, *Nonequilibrium work relations: foundations and applications*, Eur. Phys. J. B **64**, 331 (2008).

[4] U. Seifert, *Stochastic thermodynamics: principles and perspectives*, Eur. Phys. J. B **64**, 423 (2008).

[5] U. M. B. Marconi, A. Puglisi, L. Rondoni, A. Vulpiani, *Fluctuation-dissipation: Response theory in statistical physics*, Phys. Rep. **461**, 111 (2008).

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

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

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

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

[10] P. Hänggi, P. Talkner, *The other QFT*, Nature Physics **11**, 108 (2015).

[11] H. Tasaki *Jarzynski relations for quantum systems and some applications*, arXiv:cond-mat/0009244 (2000).

[12] J. Kurchan, *A quantum fluctuation theorem*, arXiv:cond-mat/0007360 (2001).

[13] S. Mukamel, *Quantum extension of the Jarzynski relation: Analogy with stochastic dephasing*, Phys. Rev. Lett. **90**, 170604 (2003).

[14] C. Jarzynski, D. K. Wójcik, *Classical and quantum fluctuation theorems for heat exchange*, Phys. Rev. Lett. **92**, 230602 (2004).

[15] P. Talkner, E. Lutz, and P. Hänggi, *Fluctuation theorems: Work is not an observable*, Phys. Rev. E **75**, 050102(R) (2007).

[16] K. Saito, Y. Utsumi *Symmetry in full counting statistics, fluctuation theorem, and relations among nonlinear transport coefficients in the presence of a magnetic field*, Phys. Rev. B **78**, 115429 (2008).

[17] H. T. Quan, H. Dong, *Quantum Crooks fluctuation theorem and quantum Jarzynski equality in the presence of a reservoir*, arXiv:0812.4955 (2008).

[18] P. Talkner, P. Hänggi, M. Morillo, *Microcanonical quantum fluctuation theorems*, Phys. Rev. E **77**, 051131 (2008).

[19] P. Talkner, M. Campisi, and P. Hänggi, *Fluctuation theorems in driven open quantum systems*, J. Stat. Mech. **P02025** (2009).

[20] D. Andrieux, P. Gaspard, T. Monnai, S. Tasaki, *The fluctuation theorem for currents in open quantum systems*, New J. Phys. **11**, 043014 (2009).

[21] M. Campisi, P. Talkner, P. Hänggi, *Fluctuation theorems for continuously monitored quantum fluxes* Phys. Rev. lett. **105**, 140601 (2010).

[22] M. Campisi, P. Talkner, P. Hänggi, *Quantum Bochkov-Kuzovlev work fluctuation theorems*, Phil. Trans. R. Soc. A **369**, 291 (2011).

[23] S. Deffner, E. Lutz, *Nonequilibrium entropy production for open quantum systems*, Phys. Rev. Lett. **107**, 140404 (2011).

[24] M. Campisi, P. Talkner, and P. Hänggi, *Influence of measurements on the statistics of work performed on a quantum system*, Phys. Rev. E **83**, 041114 (2011).

[25] P. Talkner, M. Morillo, J. Yi, P. Hänggi, *Statistics of work and fluctuation theorems for microcanonical initial states*, New J. Phys. **15**, 095001 (2013).

[26] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, Th. Beth, *Thermodynamic cost of reliability and low temperatures: Tightening Landauer's principle and the second law*, Int. J. Theor. Phys. **39**, 2717 (2000).

[27] M. Horodecki, J. Oppenheim, *Fundamental limitations for quantum and nanoscale thermodynamics*, Nat. Comm. **4**, 2059 (2013).

[28] F. Brandão, M. Horodecki, N. Ng, J. Oppenheim, S. Wehner, *The second laws of quantum thermodynamics*, PNAS **112**, 3275 (2015).

[29] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, R. W. Spekkens, *Resource theory of quantum states out of thermal equilibrium*, Phys. Rev. Lett. **111**, 250404 (2013).

[30] P. Skrzypczyk, A. J. Short, S. Popescu, *Extracting work from quantum systems*, arXiv:1302.2811 (2013).

[31] J. Åberg, *Catalytic coherence*, Phys. Rev. Lett. **113**, 150402 (2014).

[32] N. H. Y. Ng, L. Mančinska, C. Cirstoiu, J. Eisert, S. Wehner, *Limits to catalysis in quantum thermodynamics*, New. J. Phys, **17**, 085004 (2015).

[33] M. Lostaglio, D. Jennings, and T. Rudolph, *Description of quantum coherence in thermodynamics processes requires constraints beyond free energy*, Nat. Commun. **6**, 6383 (2015).

[34] M. Lostaglio, K. Korzekwa, D. Jennings, and T. Rudolph, *Quantum coherence, time-translation symmetry, and thermodynamics*, Phys. Rev. X, **5**, 021001 (2015).

[35] M. Lostaglio, M. P. Müller, M. Pastena, *Extracting work from absence of correlations*, arXiv:1409.3258 (2015).

[36] P. Ćwikliński, M. Studziński, M. Horodecki, J. Oppenheim, *Towards fully quantum second laws of thermodynamics: limitations on the evolution of quantum coherences*, arXiv:1405.5029 (2015).

[37] V. Narasimhachar, G. Gour, *Low-temperature thermodynamics with quantum coherence*, Nat. Commun. **6**, 7689 (2015).

[38] K. Korzekwa, M. Lostaglio, J. Oppenheim, D. Jennings, *The extraction of work from quantum coherence*, arXiv:1506.07875 (2015).

[39] L. Masanes, J. Oppenheim, *A derivation (and quantification) of the third law of thermodynamics*, arXiv:1412.3828 (2015).

[40] S. Wehner, M. M. Wilde, M. P. Woods, *Work and reversibility in quantum thermodynamics*, arXiv: 1506.08145 (2015).

[41] R. Gallego, J. Eisert, H. Wilming, *Defining work from operational principles*, arXiv:1504.05056 (2015).

[42] M. Woods, N. Ng, S. Wehner, *The maximum efficiency of nano heat engines depends on more than temperature*, arXiv:1506.02322 (2015).

[43] Á. M. Alhambra, J. Oppenheim, C. Perry, *What is the probability of a thermodynamical transition?*, arXiv:1504.00020 (2015).

[44] P. Skrzypczyk, A. J. Short, S. Popescu, *Work extraction and thermodynamics for individual quantum systems*, Nat. Commun. **5**, 4185 (2014).

[45] K. Kraus, *States, effects, and operations: Fundamental notions of quantum theory*, Springer lecture notes in physics, Vol. 190 (Springer, Berlin, 1983).

[46] G. E. Crooks, *Quantum operation time reversal*, Phys. Rev. A **77**, 034101 (2008).

[47] D. Petz, *Sufficient subalgebras and the relative entropy of states of a von Neumann algebra*, Commun. Math. Phys. **105**, 123 (1986).

[48] D. Petz, *Sufficiency of channels over von Neumann algebras*, Quart. J. Math. Oxford **39**, 97 (1988).

[49] H. Barnum, E. Knill, *Reversing quantum dynamics with near-optimal quantum and classical fidelity*, J. Math. Phys. **43**, 2097 (2002).

[50] P. Hayden, R. Jozsa, D. Petz, A. Winter, *Structure of states which satisfy strong subadditivity of quantum entropy with equality*, Commun. Math. Phys. **246**, 359 (2004).

[51] T. Albash, D. A. Lidar, M. Marvian, and P. Zanardi, *Fluctuation theorems for quantum processes*, Phys. Rev. E **88**, 032146 (2013).

[52] A. E. Rastegin, *Non-equilibrium equalities with unital quantum channels*, J. Stat. Mech. P06016 (2013).

[53] A. E. Rastegin, K. Życzkowski, *Jarzynski equality for quantum stochastic maps*, Phys. Rev. E **89**, 012127 (2014).

[54] J. J. Sakurai, *Modern quantum mechanics. Revised edition* (Addison-Wesley, New York, 1994).

[55] D. Andrieux, P. Gaspard, *Quantum work relations and response theory*, Phys. Rev. Lett. **100**, 230404 (2008).

[56] A. Sanpera, R. Tarrach, G. Vidal, *Quantum separability, time reversal and canonical decompositions*, arXiv: quant-ph/9707041 (1997).

[57] S. S. Bullock, G. K. Brennen, D. P. O'Leary, *Time reversal and n-qubit canonical decompositions*, J. Math. Phys. **46**, 062104 (2005).

[58] A. Peres, *Separability criterion for density matrices*, Phys. Rev. Lett. **77**, 1413 (1996).

[59] M. Horodecki, P. Horodecki, R. Horodecki, *Separability of mixed states: necessary and sufficient conditions*, Phys. Lett. A **223**, 1 (1996).

[60] I. L. Chuang, M. A. Nielsen, *Prescription for experimental determination of the dynamics of a quantum black box*, J. Mod. Opt. **44**, 2455 (1997).

[61] G. M. D'Ariano, M. De Laurentis, M. G. A. Paris, A. Porzio, S. Solimeno, *Quantum tomography as a tool for the characterization of optical devices*, J. Opt. B **4**, S127 (2002).

[62] J. B. Altepeter, D. Branning, E. Jeffrey, T. C. Wel, P. G. Kwiat, R. T. Thew, J. L. O'Brien, M. A. Nielsen, A. G. White, *Ancilla-assisted quantum process tomography*, Phys. Rev. Lett. **90**, 193601 (2003).

[63] M. Mohseni, A. T. Rezakhani, D. A. Lidar, *Quantum-process tomography: Resource analysis of different strategies*, Phys. Rev. A **77**, 032322 (2008).

[64] M. A. Nielsen, I. L. Chuang *Quantum computation and quantum information*, (Cambridge university press, Cambridge, 2000).

[65] G. N. Boschkov, Yu. E. Kuzovlev, *Nonlinear fluctuation-dissipation relations and stochastic models in nonequilibrium thermodynamics: I. Generalized fluctuation-dissipation theorem*, Physica A **106**, 443 (1981).

[66] S. Yukawa, *A quantum analogue of the Jarzynski equality*, J. Phys. Soc. Jpn. **69**, 2367 (2000).

[67] T. Monnai, S. Tasaki, *Quantum correction of fluctuation theorem*, arXiv:cond-mat/0308337 (2003).

[68] A. Engel, R. Nolte, *Jarzynski equation for a simple quantum system: Comparing two definitions of work*, Europhys. Lett. **79**, 10003 (2007).

[69] A. E. Allahverdyan, Th. M. Nieuwenhuizen, *Fluctuations of work from quantum subensambles: The case against quantum work-fluctuation theorems*, Phys. Rev. E **71**, 066102 (2005).

[70] M. F. Gelin, D. S. Kosov, *Unified approach to the derivation of work theorems for equilibrium and steady-state, classical and quantum Hamiltonian systems*, Phys. Rev. E **78**, 011116 (2008).

[71] V. Chernyak, S. Mukamel, *Effect of quantum collapse on the distribution of work in driven single molecules*, Phys. Rev. Lett. **93**, 048302 (2004).

[72] V. Chernyak, F. Šanda, S. Mukamel, *Coherence and correlations in multitime quantum measurements of stochastic quantum trajectories*, Phys. Rev. E. **73**, 036119 (2006).

[73] H. Li, J.-S. Wang, *Work distribution under continuous quantum histories*, arXiv:1304.6286 (2013).

[74] S. Mukamel, *Quantum extension of the Jarzynski relation: Analogy with stochastic dephasing* Phys. Rev. Lett. **90**, 170604 (2003).

[75] W. De Roeck, C. Maes, *Quantum version of free-energy-irreversible-work relations*, Phys. Rev. E **69**, 026115 (2004).

[76] T. Monnai, *Unified treatment of the quantum fluctuation theorem and the Jarzynski equality in terms of microscopic reversibility*, Phys. Rev. E **72**, 027102 (2005).

[77] P. Talkner, P. Hänggi, *The Tasaki-Crooks quantum fluctuation theorem*, J. Phys. A **40**, F569 (2007).

[78] A. E. Allahverdyan, *Nonequilibrium quantum fluctuations of work*, Phys. Rev. E **90**, 032137 (2014).

[79] I. Marvian, R. W. Spekkens, *Modes of asymmetry: The application of harmonic analysis to symmetric quantum dynamics and quantum reference frames*, Phys. Rev. A **90**, 062110 (2014).

[80] C. Jarzynski, *Nonequilibrium equality for free energy*

differences, Phys. Rev. Lett. **78**, 2690 (1997).

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

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

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

[84] G. Lindblad, *Non-equilibrium entropy and irreversibility* (Reidel, Dortrecht, 1983).

[85] W. H. Zurek, *Decoherence, einselection, and the quantum origins of the classical*, Rev. Mod. Phys. **75**, 715 (2003).

[86] R. Blume-Kohout, *Quantum Darwinism: Entanglement, branches, and the emergent classicality of redundantly stored quantum information*, Phys. Rev. A **73**, 062310 (2006).

[87] W. H. Zurek, *Quantum Darwinism*, Nature Phys. **5**, 181 (2009).

[88] B. Cleuren, C. Van den Broeck, R. Kawai, *Fluctuation and dissipation of work in a Joule experiment*, Phys. Rev. Lett. **96**, 050601 (2006).

[89] M. Campisi, *Complementary expressions for the entropy-from-work theorem*, Phys. Rev. E **78**, 012102 (2008).

[90] E. B. Davies, *Quantum theory of open systems* (Academic Press, NewYork, 1976).

[91] A. S. Holevo, *Probabilistic and statistical aspects of quantum theory* (North Holland, Amsterdam,1982).

[92] A. S. L. Malabarba, A. J. Short, P. Kammerlander, *Clock-driven quantum thermal engines*, New. J. Phys. **17**, 045027 (2015).

[93] M. F. Frenzel, D. Jennings, T. Rudolph, *Autonomous quantum thermal machines and quantum to classical energy flow*, arXiv:1508.02720 (2015).

[94] H.-P. Breuer, F. Petruccione, *The theory of open quantum systems* (Oxford University Press, Oxford, 2010).

[95] R. Chetrite, K. Mallick, *Quantum fluctuation relations for the Lindblad master equation*, J. Stat. Phys. **148**, 480 (2012).

[96] B. Leggio, A. Napoli, H.-P. Breuer, A. Messina, *Fluctuation theorems for non-Markovian quantum processes*, arXiv:1301.1553 (2013).

[97] B. Leggio, A. Napoli, A. Messina, H.-P. Breuer, *Entropy production and information fluctuations along quantum trajectories*, arXiv:1305.6733 (2013).

[98] Z.-W. Bai, *Quantum work fluctuation theorem: Nonergodic Brownian motion case*, Phys. Lett. A **378**, 1679 (2014).

[99] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco Jr., C. Bustamante, *Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality*, *Science* **296**, 1832 (2002).

[100] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco Jr., C. Bustamante, *Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies*, Nature **437**, 231 (2005).

[101] F. Douarche, S. Ciliberto, A. Petrosyan, I. Rabbiosi, *An experimental test of the Jarzynski equality in a mechanical experiment*, Europhys. Lett. **70**, 593 (2005).

[102] N. C. Harris, Y. Song, C.-H. Kiang, *Experimental free energy surface reconstruction from single-molecule force spectroscopy using Jarzynski's equality*, Phys. Rev. Lett. **99**, 068101 (2007).

[103] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano, *Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality*, Nature Physics, **6**, 988 (2010).

[104] J. Gieseler, R. Quidant, C. Dellago, L. Novotny *Dynamic relaxation of a levitated nanoparticle from a non-equilibrium steady state*, Nature Nanotechnology **9**, 358 (2014).

[105] G. Huber, F. Schmidt-Kaler, S. Deffner, E. Lutz, *Employing trapped cold ions to verify the quantum Jarzynski equality*, Phys. Rev. Lett. **101**, 070403 (2008).

[106] M. Heyl, S. Kehrein, *Crooks relation in optical spectra: Universality in work distributions for weak local quenches*, Phys. Rev. Lett. **108**, 190601 (2012).

[107] J. P. Pekola, P. Solinas, A. Shnirman, D. V. Averin, *Calorimetric measurement of work in a quantum system*, New J. Phys. **15**, 115006 (2013).

[108] R. Dorner, S. R. Clark, L. Heaney, R. Fazio, J. Goold, V. Vedral, *Extracting quantum work statistics and fluctuation theorems by single-qubit interferometry*, Phys. Rev. Lett. **110**, 230601 (2013).

[109] L. Mazzola, G. De Chiara, M. Paternostro, *Measuring the characteristic function of the work distribution*, Phys. Rev. Lett. **110**, 230602 (2013).

[110] M. Campisi, R. Blattmann, S. Kohler, D. Zueco, P. Hänggi, *Employing circuit QED to measure non-equilibrium work fluctuations*, New J. Phys. **15**, 105028 (2013).

[111] M. Campisi, J. Pekola, R. Fazio, *Nonequilibrium fluctuations in quantum heat engines: Theory, example, and possible solid state experiments*, arXiv: 1412.0898 (2014).

[112] T. B. Batalhão, A. M. Souza, L. Mazzola, R. Auccaise, R. S. Sarthour, I. S. Oliveira, J. Go old, G. De Chiara, M. Paternostro, R. M. Serra, *Experimental reconstruction of work distribution and study of fluctuation relations in a closed quantum system*, Phys. Rev. Lett. **113**, 140601 (2014).

[113] S. An, J.-N. Zhang, M. Um, D. Lv, Y. Lu, J. Zhang, Z.-Q. Yin, H. T. Quan, K. Kim, *Experimental test of the quantum Jarzynski equality with a trapped-ion system*, Nature Physics **11**, 193 (2015).

[114] S. Schuler, T. Speck, C. Tietz, J. Wrachtrup, U. Seifert, *Experimental test of the fluctuation theorem for a driven two-level system with time-dependent rates*, Phys. Rev. Lett. **94**, 180602 (2005).

[115] S. Popescu, A. J. Short, A. Winter, *Entanglement and the foundations of statistical mechanics*, Nature Physics, **2**, 754 (2006).

[116] S. Goldstein, J. L. Lebowitz, R. Tumulka, N. Zanghì, *Canonical typicality*, Phys. Rev. Lett **96**, 050403 (2006).

[117] J. Yi, Y. W. Kim, P. Talkner, *Work fluctuations for Bose particles in grand canonical initial states*, Phys. Rev. E **85**, 051107 (2012).

[118] T. Sagawa, M. Ueda, *Generalized Jarzynski equality under nonequilibrium feedback control*, Phys. Rev. Lett. **104**, 090602 (2010).

[119] Y. Morikuni, H. Tasaki, *Quantum Jarzynski-Sagawa-Ueda relations*, J. Stat. Phys. **143**, 1 (2011).

[120] M. Horodecki, J. Oppenheim, *(Quantumness in the context of) resource theories*, Int. J. Mod. Phys. B **27**, 1345019 (2013).

[121] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, N. Y. Halpern, *The resource theory of informational nonequilibrium in thermodynamics*, Phys. Rep. **583**, 1 (2015).

[122] O. C. O. Dahlsten, R. Renner, E. Rieper, V. Vedral, *Inadequacy of von Neumann entropy for characterizing extractable work*, New J. Phys. **13**, 053015 (2011).

[123] L. del Rio, J. Åberg, R. Renner, O. Dahlsten, V. Vedral, *The thermodynamic meaning of negative entropy*, Nature **474**, 61 (2011).

[124] J. Åberg, *Truly work-like work extraction via a single-shot analysis*, Nat. Commun. **4**, 1925 (2013).

[125] D. Egloff, O. C. O. Dahlsten, R. Renner, V. Vedral, *A measure of majorization emerging from single-shot statistical mechanics*, New J. Phys. **17**, 073001 (2015).

[126] D. Egloff, *Work value of mixed states in single instance work extraction games*, Master's thesis, ETH Zurich (2010).

[127] P. Faist, F. Dupuis, J. Oppenheim, R. Renner, *The minimal work cost of information processing*, Nat. Commun. **6**, 7669 (2015).

[128] D. Reeb, M. M. Wolf, *An improved Landauer principle with finite-size corrections*, New J. Phys. **16**, 103011 (2014).

[129] J. M. Renes, *Work cost of thermal operations in quantum and nano thermodynamics*, arXiv:1402.3496 (2014).

[130] N, Y. Halpern, J. M. Renes, *Beyond heat baths: Generalized resource theories for small-scale thermodynamics*, arXiv:1409.3998 (2014).

[131] P. Kammerlander, J. Anders, *Coherence and measurement in quantum thermodynamics*, arXiv:15032.0267 (2015).

[132] J. Gemmer, J. Anders, *From single-shot towards general work extraction in a quantum thermodynamic framework*, New J. Phys. **17**, 085006 (2015).

[133] N. Y. Halpern, A. J. P. Garner, O. C. O. Dahlsten, V. Vedral, *Introducing one-shot work into fluctuation relations*, New J. Phys. **17**, 095003 (2015).

[134] S. Salek, K. Wiesner, *Fluctuations in single-shot ϵ-deterministic work extraction*, arXiv:1504.05111 (2015).

[135] O. C. O. Dahlsten, M.-S. Choi, D. Braun, A. J. P. Garner, N. Y. Halpern, V. Vedral, *Equality for worst-case work at any protocol speed*, arXiv:1504.05152 (2015).

[136] N. Y. Halpern, A. J. P. Garner, O. C. O. Dahlsten, V. Vedral, *What's the worst that could happen? One-shot dissipated work from Rényi divergences*, arXiv:1505.06217 (2015).

[137] J. Goold, M. Huber, A. Riera, L. del Rio, P. Skrzypczyk, *The role of quantum information in thermodynamics - a topical review*, arXiv:1505.0783 (2015).

[138] S. Vinjanampathy, J. Anders, *Quantum thermodynamics*, arXiv:1508.06099 (2015).

[139] T. Sagawa, M. Ueda, *Second law of thermodynamics with discrete quantum feedback control*, Phys. Rev. Lett. **100**, 080403 (2008).

[140] E. Kreyszig, *Introductory functional analysis with applications* (Wiley, 1989).

[141] R. A. Horn, C. R. Johnson, *Matrix analysis*, 2nd Ed. (Cambridge University Press, Cambridge, 2013).

[142] R. Bhatia, *Matrix analysis*, Graduate texts in mathematics 169 (Springer, New York, 1997).

[143] G. E. Crooks, *Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems*, J. Stat. Phys. **90**, 1481 (1998).

[144] A. O. Caldeira, A. J. Legget, *Path integral approach to quantum Brownian motion*, Physica A **121**, 587 (1983).