



PAPER • OPEN ACCESS

Quantum nonequilibrium equalities with absolute irreversibility

To cite this article: Ken Funo *et al* 2015 *New J. Phys.* **17** 075005

View the [article online](#) for updates and enhancements.

You may also like

- [Dynamics of skew products of interval maps](#)
L. S. Efremova
- [Sobolev-orthogonal systems of functions and some of their applications](#)
I. I. Sharapudinov
- [Schur's criterion for formal power series](#)
V. I. Buslaev



PAPER

Quantum nonequilibrium equalities with absolute irreversibility

Ken Funo¹, Yûto Murashita¹ and Masahito Ueda^{1,2}¹ Department of Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan² RIKEN Center for Emergent Matter Science (CEMS), Wako 351-0198, Japan

E-mail: funo@cat.phys.s.u-tokyo.ac.jp

Keywords: fluctuation theorems, information thermodynamics, feedback control, free expansion

RECEIVED

18 December 2014

REVISED

8 May 2015

ACCEPTED FOR PUBLICATION

29 May 2015

PUBLISHED

9 July 2015

Content from this work
may be used under the
terms of the [Creative
Commons Attribution 3.0
licence](#).

Any further distribution of
this work must maintain
attribution to the
author(s) and the title of
the work, journal citation
and DOI.



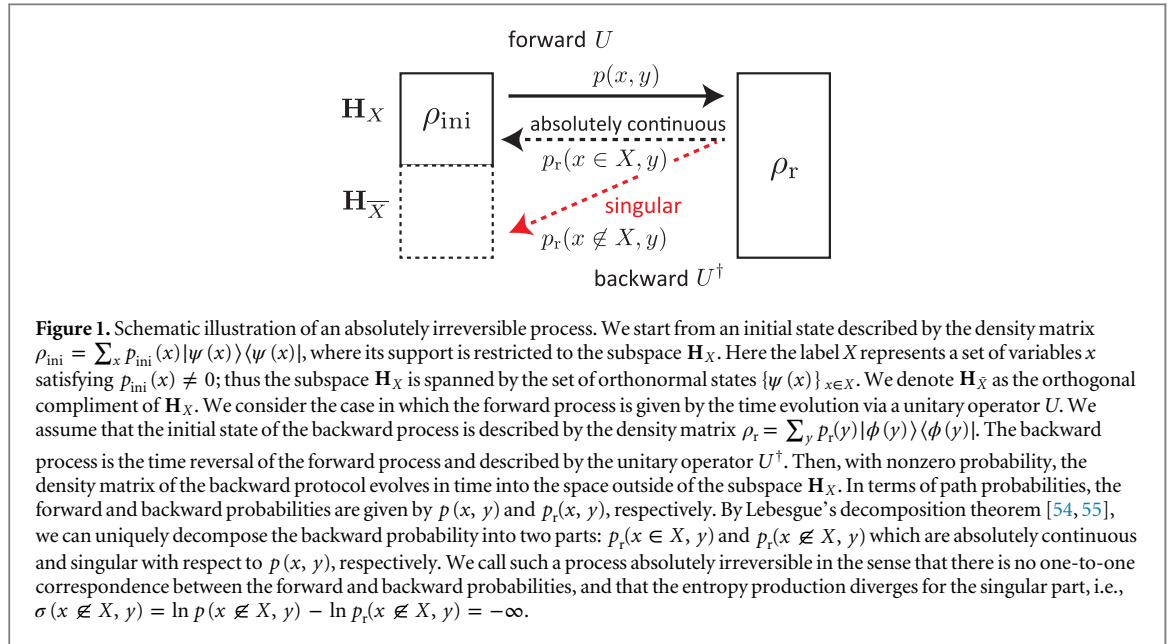
Abstract

We derive quantum nonequilibrium equalities in absolutely irreversible processes. Here by absolute irreversibility we mean that in the backward process the density matrix does not return to the subspace spanned by those eigenvectors that have nonzero weight in the initial density matrix. Since the initial state of a memory and the postmeasurement state of the system are usually restricted to a subspace, absolute irreversibility occurs during the measurement and feedback processes. An additional entropy produced in absolutely irreversible processes needs to be taken into account to derive nonequilibrium equalities. We discuss a model of a feedback control on a qubit system to illustrate the obtained equalities. By introducing N heat baths each composed of a qubit and letting them interact with the system, we show how the entropy reduction via feedback control can be converted into work. An explicit form of extractable work in the presence of absolute irreversibility is given.

1. Introduction

Nonequilibrium equalities [1–27] such as fluctuation theorems and Jarzynski equalities have attracted a great deal of interest in the field of nonequilibrium statistical mechanics. They give general insights into thermodynamic quantities in nonequilibrium processes irrespective of details of individual systems. For example, the Jarzynski equality relates work done on the system in a nonequilibrium process to the equilibrium free-energy difference. Nonequilibrium equalities have been obtained in both classical [1–12] and quantum [13–27] systems, and generalized to situations involving feedback control [28–32]. Due to recent advancement in experimental techniques, nonequilibrium equalities have been experimentally verified for classical systems such as a single-molecule RNA [33, 34], and they have been vindicated in a quantum regime using a trapped ion system [35]. Also, feedback control on a Brownian particle was carried out to experimentally demonstrate Maxwell's demon [36], and the generalized Jarzynski equality for a feedback-controlled system was verified [37].

It is known that the Jarzynski equality is inapplicable to such cases as free expansion [38–42] and feedback control involving projective measurements [31] because in these cases there exist those forward paths with vanishing probability that have the corresponding backward paths with nonzero probabilities. We shall call such processes absolutely irreversible. We give examples of absolutely irreversible processes in section 4. Recently, nonequilibrium equalities have been obtained that can be applied to absolutely irreversible processes [45, 46]. We extend this idea to quantum systems and derive quantum fluctuation theorems and Jarzynski equalities with absolute irreversibility. For the quantum case, absolute irreversibility occurs when the initial state $\rho_{\text{ini}} = \sum_x p_{\text{ini}}(x) |\psi(x)\rangle \langle \psi(x)|$ (with $p_{\text{ini}}(x) \neq 0$ for $x \in X$) is restricted to a subspace \mathbf{H}_X (spanned by $\{|\psi(x)\rangle\}_{x \in X}$) of the total Hilbert space, and the density matrix of the backward process is not confined to that subspace, i.e., $\lambda = \sum_x \langle \psi(x) | \bar{\rho} | \psi(x) \rangle \neq 0$, where $\bar{\rho}$ is the final density matrix of the backward process (see figure 1). Then, the initially localized state expands into a larger space, as in free expansion, with the probability λ . Absolute irreversibility is likely to occur in measurement and feedback processes since the initial state of the memory and the postmeasurement state are localized in general, and the projective measurement on the memory and the effect of (inefficient) feedback control let these states expand to a space larger than the subspace of the initial state, resulting in additional entropy production. By subtracting the absolutely irreversible part in a



mathematically well-defined manner, we derive those nonequilibrium equalities for measurement and feedback processes that give stronger restrictions on entropy productions or work compared with previously known results [32, 47, 48]. We note that a quantum Jarzynski equality under feedback control with projective measurement was obtained in [31], where the issue of absolute irreversibility was circumvented by introducing classical errors on measurement outcomes.

This paper is organized as follows. In section 2, we derive nonequilibrium equalities without feedback control for quantum systems. We introduce the concept of absolute irreversibility and discuss how the nonequilibrium equalities are modified by this effect. In section 3, we derive nonequilibrium equalities with feedback control in the presence of absolute irreversibility during feedback control and the measurement process. In section 4, we give an example of the quantum piston to analyze the free expansion of the gas with absolute irreversibility. We also give an example of the feedback control on a qubit system to illustrate our work. In section 5, we summarize the main results of this paper.

2. Nonequilibrium equalities in the absence of feedback control

2.1. Setup

Let the initial state of the system be ρ_{ini} and let the system evolve in time according to a unitary evolution:

$$U = \text{Texp} \left(-i \int_0^{t_{\text{fin}}} H(t) dt \right), \quad (1)$$

where $H(t)$ is the time-dependent Hamiltonian and T is the time-ordering operator. The final state of the system is given by

$$\rho_{\text{fin}} = U \rho_{\text{ini}} U^\dagger. \quad (2)$$

We define the general form of the entropy production, which measures the irreversibility of the process, as

$$\langle \sigma \rangle = -S(\rho_{\text{ini}}) - \text{Tr}[\rho_{\text{fin}} \ln \rho_r], \quad (3)$$

where $S(\rho_{\text{ini}}) = -\text{Tr}[\rho_{\text{ini}} \ln \rho_{\text{ini}}]$ is the von Neumann entropy and ρ_r is a reference state which can be chosen arbitrarily [49]. Because of equation (2), the entropy production defined here gives the quantum relative entropy between the final state and the reference state:

$$\langle \sigma \rangle = -S(\rho_{\text{fin}}) - \text{Tr}[\rho_{\text{fin}} \ln \rho_r] = S(\rho_{\text{fin}} \| \rho_r) \geq 0, \quad (4)$$

where the inequality results from the nonnegativity of the quantum relative entropy [53]. The entropy production in equation (3) is defined as the quantum analog of the classical one given by

$$\sigma[\Gamma] = \ln \frac{P[\Gamma]}{\tilde{P}[\Gamma^\dagger]}, \quad (5)$$

and its average is given by

$$\langle \sigma \rangle = \int d\Gamma P[\Gamma] \sigma[\Gamma] = \int d\Gamma P[\Gamma] \ln \frac{P[\Gamma]}{\tilde{P}[\Gamma^\dagger]}. \quad (6)$$

Here Γ denotes the trajectory of the system in phase space and Γ^\dagger is the trajectory of the reference dynamics, and $P[\cdot]$ and $\tilde{P}[\cdot]$ are the forward and reference probability distributions. The relation between choices of the reference probability and specific meanings of the entropy production is explained in [8] for classical systems. For example, if we take the reference dynamics as the time reversal and the reference state as the final state of the forward process, equation (5) reduces to the total entropy production [6]. If we take the reference dynamics as the dual-reversed process, in which both time and steady flow are reversed, and the reference state as the steady state, equation (5) is related to the excess heat [11]. Note that the fluctuation theorem was tested through measurements of the total entropy production [9, 10] and the excess heat [12]. For quantum systems, there are at least two possible choices for the reference state which lead to the dissipated work and the total entropy production, as we show in the next section. Thus, our formulation based on the reference density matrix offers a unified approach to deriving various fluctuation theorems.

2.1.1. Examples of the choice of reference states and the corresponding entropy productions

(1) *Dissipated work.* We define the dissipated work W_d in terms of the work $\langle W \rangle$ done by the system (or work that can be extracted from the system)³ and the equilibrium free energy difference $\Delta F = F_{\text{fin}} - F_{\text{ini}}$ as

$$W_d = -\beta \langle W \rangle - \beta \Delta F, \quad (7)$$

and choose the initial state to be the canonical distribution

$$\rho_{\text{ini}} = e^{-\beta (H_{\text{ini}} - F_{\text{ini}})}, \quad (8)$$

where $H_{\text{ini}} = H(0)$. If we choose the reference state to be the canonical distribution with respect to the final Hamiltonian $H_{\text{fin}} = H(t_{\text{fin}})$ as

$$\rho_r = e^{-\beta (H_{\text{fin}} - F_{\text{fin}})}, \quad (9)$$

then equation (3) reduces to the dissipated work

$$\langle \sigma \rangle = -\beta (\langle W \rangle + \Delta F) = W_d \geq 0, \quad (10)$$

where we define work during the nonequilibrium process by the energy change of the system:

$$\langle W \rangle = \text{Tr}[H_{\text{ini}} \rho_{\text{ini}}] - \text{Tr}[H_{\text{fin}} \rho_{\text{fin}}]. \quad (11)$$

The above argument applies to an isolated system. In the presence of a heat bath, the total Hamiltonian in equation (1) is given by

$$H(t) = H^S(t) + V^{SB}(t) + H^B, \quad (12)$$

where $H^S(t)$ is the system Hamiltonian which depends on time via external control parameters, H^B is the Hamiltonian of the heat bath, and $V^{SB}(t)$ is the interaction between the system and the heat bath. In equation (12), we assume that the interaction $V^{SB}(t)$ is either turned off at the initial and final states, i.e.

$$V^{SB}(0) = V^{SB}(t_{\text{fin}}) = 0, \quad (13)$$

or the interaction is independent of time and very weak, i.e.

$$V^{SB}(t) = \kappa V^{SB}, \quad \kappa \ll 1. \quad (14)$$

Later, we discuss the validity of these assumptions. We also use the abbreviations $H_{\text{ini}}^S = H^S(0)$ and $H_{\text{fin}}^S = H^S(t_{\text{fin}})$. We first consider the case (13). Then

$$\rho_{\text{ini}} = e^{-\beta (H_{\text{ini}}^S - F_{\text{ini}}^S)} \otimes e^{-\beta (H^B - F^B)} \quad (15)$$

holds and the choice of reference state in equation (9) leads to

$$\rho_r = e^{-\beta (H_{\text{fin}}^S - F_{\text{fin}}^S)} \otimes e^{-\beta (H^B - F^B)}. \quad (16)$$

Combining equations (3), (15) and (16), we reproduce equation (10):

$$\langle \sigma \rangle = -\beta (\langle W \rangle + \Delta F^S) = W_d \geq 0, \quad (17)$$

where the work appearing in equation (17) is given by

$$\langle W \rangle = \text{Tr}[H_{\text{ini}}^S \rho_{\text{ini}}^S] - \text{Tr}[H_{\text{fin}}^S \rho_{\text{fin}}^S] + Q. \quad (18)$$

³ We keep this sign convention throughout this paper.

Here, we interpret the heat Q as the energy that is transferred from the heat bath to the system during the process:

$$Q = \text{Tr}[H^B \rho_{\text{ini}}^B] - \text{Tr}[H^B \rho_{\text{fin}}^B]. \quad (19)$$

Now let us consider the validity of the assumption (13) we made for the interaction $V(t)$. We have in mind a system that is attached to the heat bath at the initial time and detached at the final time. One example of a system satisfying this condition is a cavity field interacting with a sequence of atoms passing through the cavity, where atoms can act as a heat bath to the cavity field [50]. There are some subtlety for the definition of work in this case because there might be a contribution to work (18) from the action of switching on and off the interaction. We can avoid this subtlety by assuming that the interaction is turned on and off adiabatically. We also note that we can adopt the framework of continuous measurement (by considering many heat baths interacting with a system one by one) and by taking the limit in which the interaction time with each environment is sufficiently small, but the coupling strength is assumed to scale as the inverse of the square root of the interaction time. In this limit, the stochastic master equation was derived, and fluctuation theorems have been obtained [23].

Next, let us consider the case where we assume equation (14) for the interaction $V(t)$. We note that when the interaction is always present, the initial state has a correlation between the system and the heat bath. However, in the weak-coupling limit, equation (15) is correct up to the second order of the coupling strength:

$$\rho_{\text{ini}} = e^{-\beta(H_{\text{ini}} - F_{\text{ini}})} = e^{-\beta(H_{\text{ini}}^S - F_{\text{ini}}^S)} \otimes e^{-\beta(H^B - F^B)} + O(\kappa^2). \quad (20)$$

When we use the definition of work and heat as given in equations (18) and (19), we must assume that the energy change due to the interaction energy is small. We assume that the total energy change is divided into the energy change of the system and that of the heat bath for weak coupling as discussed in [26]:

$$\text{Tr}[H_{\text{ini}} \rho_{\text{ini}}] - \text{Tr}[H_{\text{fin}} \rho_{\text{fin}}] \simeq \left(\text{Tr}[H_{\text{ini}}^S \rho_{\text{ini}}^S] - \text{Tr}[H_{\text{fin}}^S \rho_{\text{fin}}^S] \right) + \left(\text{Tr}[H^B \rho_{\text{ini}}^B] - \text{Tr}[H^B \rho_{\text{fin}}^B] \right). \quad (21)$$

The main results for the rest of this paper are based on the assumption (13) for the interaction, but the same result can also be derived if we assume (14) instead of (13) (in particular, a quantum Jarzynski equality without absolute irreversibility was derived in [26] for the weak-coupling interaction).

(2) *Total entropy production.* To relate the entropy production to the total entropy production, we consider a system composed of a system and a heat bath, and use the same Hamiltonian as in equation (12). We assume that the initial state of the heat bath is given by the canonical distribution

$$\rho_{\text{ini}}^{SB} = \rho_{\text{ini}}^S \otimes e^{-\beta(H^B - F^B)}, \quad (22)$$

and choose the reference state as follows:

$$\rho_r^{SB} = \rho_{\text{fin}}^S \otimes e^{-\beta(H^B - F^B)}. \quad (23)$$

Combining equations (3), (22) and (23), we obtain

$$\langle \sigma \rangle = \Delta S - \beta Q = \sigma_{\text{tot}} \geq 0, \quad (24)$$

where $\Delta S = S(\rho_{\text{fin}}^S) - S(\rho_{\text{ini}}^S)$ is a change in the von Neumann entropy of the system and Q is the heat defined in equation (19). If we interpret heat as the entropy produced in the heat bath, equation (24) expresses the entropy that is produced for the total system during the protocol; σ_{tot} is therefore called the total entropy production.

Equation (4) leads to second-law-like inequalities for entropy productions, e.g., for dissipated work and total entropy production, and the nonnegativity of the entropy production shows that there is dissipation in a given process [49]. The process is thermodynamically reversible if and only if the equality in (4) holds (for example, if the dissipated work W_d or the total entropy production σ_{tot} vanishes).

2.2. Quantum fluctuation theorem

Next, we derive quantum fluctuation theorems by expressing the initial state in the diagonal basis as

$\rho_{\text{ini}} = \sum_x p_{\text{ini}}(x) |\psi(x)\rangle \langle \psi(x)|$, where $\{|\psi(x)\rangle\}$ is an orthonormal basis set, and the reference state as $\rho_r = \sum_y p_r(y) |\phi(y)\rangle \langle \phi(y)|$. The entropy production can then be calculated as

$$\begin{aligned}
\langle \sigma \rangle &= \sum_x p_{\text{ini}}(x) \ln p_{\text{ini}}(x) - \text{Tr} \left[\rho_{\text{fin}} \ln \left(\sum_y p_r(y) |\phi(y)\rangle \langle \phi(y)| \right) \right] \\
&= \sum_x p_{\text{ini}}(x) \ln p_{\text{ini}}(x) - \sum_y \langle \phi(y) | \rho_{\text{fin}} | \phi(y) \rangle \ln p_r(y) \\
&= \sum_x p_{\text{ini}}(x) \ln p_{\text{ini}}(x) - \sum_{x,y} p(x, y) \ln p_r(y) \\
&= \sum_{x,y} p(x, y) \ln \frac{p_{\text{ini}}(x)}{p_r(y)}, \tag{25}
\end{aligned}$$

where

$$p(x, y) = p_{\text{ini}}(x) p(y|x) \tag{26}$$

and

$$p(y|x) = |\langle \phi(y) | U | \psi(x) \rangle|^2 \tag{27}$$

is the transition probability from the state $|\psi(x)\rangle$ to $|\phi(y)\rangle$ via a unitary operator U . Such a transition is characterized by a set of labels (x, y) . In deriving the third line in equation (25), we used the relation

$$\begin{aligned}
\langle \phi(y) | \rho_{\text{fin}} | \phi(y) \rangle &= \langle \phi(y) | U \rho_{\text{ini}} U^\dagger | \phi(y) \rangle \\
&= \sum_x p_{\text{ini}}(x) |\langle \phi(y) | U | \psi(x) \rangle|^2. \tag{28}
\end{aligned}$$

From equation (25), we define the following unaveraged entropy production:

$$\sigma(x, y) = \ln \frac{p_{\text{ini}}(x)}{p_r(y)}. \tag{29}$$

Next, we introduce the reference probability distribution

$$p_r(x, y) = p_r(y) \tilde{p}(x|y), \tag{30}$$

where

$$\tilde{p}(x|y) = |\langle \psi(x) | U^\dagger | \phi(y) \rangle|^2 = p(y|x) \tag{31}$$

is the transition probability from $|\phi(y)\rangle$ to $|\psi(x)\rangle$ via U^\dagger . Equation (30) gives the probability of the backward process that starts from the reference state and evolves in time via U^\dagger . It follows from equation (31) that the entropy production is expressed in terms of the forward and reference probabilities as follows:

$$\sigma(x, y) = \ln \frac{p_{\text{ini}}(x) p(y|x)}{p_r(y) \tilde{p}(x|y)} = \ln \frac{p(x, y)}{p_r(x, y)}. \tag{32}$$

Now we derive the quantum fluctuation theorem by using the above definition of entropy production (29). Since the reference probability is normalized to unity, we have

$$\sum_{x,y} p_r(x, y) = \text{Tr} [U^\dagger \rho_r U] = \text{Tr} [\rho_r] = 1. \tag{33}$$

The entropy production is given by the ratio between the forward and reference probabilities. However, if the forward probability vanishes and the corresponding reference probability does not, the logarithm of the ratio in equation (32) diverges. To deal with such singular situations, we divide the reference probability into two parts:

$$1 = \sum_{x,y} p_r(x, y) = \sum_{x \in X, y} p_r(x, y) + \sum_{x \notin X, y} p_r(x, y), \tag{34}$$

where $X = \{x | p_i(x) \neq 0\}$. Since we can take the ratio between the forward and reference probabilities for the first term of the right-hand side of equation (34), we have

$$\begin{aligned}
1 &= \sum_{x \in X, y} \frac{p_r(x, y)}{p_r(x, y)} p(x, y) + \lambda \\
&= \sum_{x \in X, y} p(x, y) e^{-\sigma(x, y)} + \lambda \\
&= \langle e^{-\sigma} \rangle + \lambda, \tag{35}
\end{aligned}$$

where

$$\lambda = \sum_{x \notin X, y} p_t(x, y) \quad (36)$$

gives the total probability of those backward processes that do not return to the subspace spanned by $\{|\psi(x)\rangle\}_{x \in X}$. In an ordinary irreversible process, the process is stochastically reversible in the sense that the backward path returns to the initial state with a nonzero probability since there is a one-to-one correspondence between the forward and backward paths, i.e., the entropy production is finite⁴ for all (x, y) in equation (32). However, the path labeled by the set of variables $(x \notin X, y)$ is not even stochastically reversible since the formal definition of the entropy production negatively diverges, i.e.

$$\sigma(x \notin X, y) = \ln \frac{0}{p_t(x, y)} = -\infty, \quad (37)$$

and we call this type of irreversibility absolute irreversibility [46]. A schematic illustration of an absolutely irreversible process is shown in figure 1.

By rewriting equation (35), we obtain a quantum fluctuation theorem in the presence of absolute irreversibility:

$$\langle e^{-\sigma} \rangle = 1 - \lambda. \quad (38)$$

By using the Jensen's inequality, i.e., $\langle e^x \rangle \geq e^{\langle x \rangle}$, we obtain the following inequality for the entropy production:

$$\langle \sigma \rangle \geq -\ln(1 - \lambda). \quad (39)$$

This result shows that in the presence of absolute irreversibility the entropy production must be positive and not less than $-\ln(1 - \lambda) > 0$, giving a stronger constraint compared with the second law-like inequality $\langle \sigma \rangle \geq 0$. Note that only when there is no absolute irreversibility, i.e., $\lambda = 0$, the conventional fluctuation theorem is reproduced: $\langle e^{-\sigma} \rangle = 1$.

In the classical case, a decomposition similar to equation (34) can be carried out in a general framework using the probability measure [46]. To see this, let us denote the forward and reference probability measures in phase space as \mathcal{M} and \mathcal{M}^r , respectively. According to Lebesgue's decomposition theorem [54, 55], \mathcal{M}^r is uniquely decomposed into two parts: $\mathcal{M}^r = \mathcal{M}_{AC}^r + \mathcal{M}_S^r$, where \mathcal{M}_{AC}^r and \mathcal{M}_S^r are absolutely continuous and singular with respect to \mathcal{M} , respectively. Provided that the probability distribution of a quantum process in this setup is labeled by discrete variables, the decomposition of the reference probability is carried out by dividing variables into two parts: the variables with nonvanishing forward probabilities ($x \in X$) and the variables with vanishing forward probabilities ($x \notin X$). Then, \mathcal{M}_{AC}^r corresponds to $p^r(x \in X, y)$ and \mathcal{M}_S^r corresponds to $p^r(x \notin X, y)$. The uniqueness of this decomposition is ensured by Lebesgue's decomposition theorem.

Note that the absence of absolute irreversibility and the requirement for the 'ergodic consistency' discussed in [7] are different concepts. The ergodic consistency requires that for all initial phase space $\Gamma(0)$ with nonzero probability $f(\Gamma(0), 0) \neq 0$, the corresponding initial probability distribution of the time-reversed process is nonzero, i.e., $f(\Gamma^\dagger(t), 0) \neq 0$, where $f(\Gamma, s)$ is the probability distribution of the system in the phase space point Γ at time s . Therefore, the ergodic consistency requires that for all nonzero forward path probabilities, the corresponding reference (backward) path probabilities are nonzero. In contrast, the absence of absolute irreversibility requires that for all nonzero reference (backward) path probabilities, the corresponding forward path probabilities are nonzero. The two conditions are different in the sense that in the former case, the backward protocol (especially, the initial state of the backward process) is fixed and thus the condition on the forward path probability is imposed, whereas in the latter case, the forward protocol is fixed and thus the condition on the backward path probability is imposed.

2.3. Quantum Jarzynski equality

We now derive the quantum Jarzynski equality by assuming that the initial state is given by the canonical distribution (15) and by taking the reference state as given in equation (16). For convenience, we use the notation $x = (x_1, x_2)$ and $y = (y_1, y_2)$, where the subscript 1 refers to the system and 2 to the heat bath. By assumption, we have

$$|\psi(x)\rangle = |E_{\text{ini}}^S(x_1)\rangle \otimes |E^B(x_2)\rangle, \quad |\phi(y)\rangle = |E_{\text{fin}}^S(y_1)\rangle \otimes |E^B(y_2)\rangle, \quad (40)$$

where $|E_{\text{ini}}^S(x_1)\rangle$ and $|E_{\text{fin}}^S(y_1)\rangle$ are energy eigenstates of the initial and final Hamiltonians of the system, respectively, and $|E^B(x_2)\rangle$ is the energy eigenstate of the heat bath. Now the unaveraged entropy production (29) is related to work by

⁴ To be precise, when $p_t(x, y) = 0 \wedge p(x, y) \neq 0$, the entropy production in equation (32) positively diverges. However, this does not cause any problem because $e^{-\sigma(x, y)} = p_t(x, y)/p(x, y) = 0$ remains finite and so is the first term on the left-hand side of equation (35).

$$\sigma(x, y) = -\beta(W(x, y) + \Delta F^S), \quad (41)$$

where

$$\begin{aligned} W(x, y) &= E_{\text{ini}}^S(x_1) - E_{\text{fin}}^S(y_1) + E^B(x_2) - E^B(y_2) \\ &= E_{\text{ini}}^S(x_1) - E_{\text{fin}}^S(y_1) + \beta Q(x_2, y_2) \end{aligned} \quad (42)$$

is the unaveraged work done by the system and $Q(x_2, y_2)$ is the unaveraged heat that flows into the system.

Substituting equation (41) into equation (38), we obtain the quantum Jarzynski equality in the presence of absolute irreversibility:

$$\langle e^{\beta(W+\Delta F)} \rangle = 1 - \lambda. \quad (43)$$

Substituting equation (41) into equation (39), we obtain the second-law like inequality

$$\langle W \rangle \leq -\Delta F^S + k_B T \ln(1 - \lambda). \quad (44)$$

Since the canonical distribution is full rank, i.e.

$$p_{\text{ini}}^S(x_1) = e^{-\beta(E_{\text{ini}}^S(x_1) - F_{\text{ini}}^S)} \neq 0 \quad \text{for all } x_1, \quad (45)$$

there is no absolute irreversibility, i.e., $\lambda = 0$. However, if we prepare the initial state in a local equilibrium state, there is a possibility that the process is absolutely irreversible and the effect of nonzero λ restricts the extractable work. For simplicity, let us divide the Hamiltonian of the system into two parts $H_{\text{ini}}^S = H_{\text{ini}}^{S_X} \oplus H_{\text{ini}}^{S_{\bar{X}}}$ and prepare the initial state as the canonical distribution that is restricted to the subspace corresponding to $H_{\text{ini}}^{S_X}$:

$$\rho_{\text{ini}}^S = e^{-\beta(H_{\text{ini}}^{S_X} - F_{\text{ini}}^{S_X})} = \sum_{x_1 \in X} e^{-\beta(E_{\text{ini}}^S(x_1) - F_{\text{ini}}^{S_X})} |E_{\text{ini}}^S(x_1)\rangle \langle E_{\text{ini}}^S(x_1)|, \quad (46)$$

where $\{|E_{\text{ini}}^S(x_1)\rangle\}_{x_1 \in X}$ is the energy eigenstate of the Hamiltonian $H_{\text{ini}}^{S_X}$. Then, $\{|E_{\text{ini}}^S(x_1)\rangle\}_{x_1 \notin X}$ is the energy eigenstate of the Hamiltonian $H_{\text{ini}}^{S_{\bar{X}}}$. Now λ is given by the total probability of those backward processes in which the system returns to the subspace spanned by $\{|E_{\text{ini}}^S(x_1)\rangle\}_{x_1 \notin X}$:

$$\lambda = \sum_{x_1 \notin X} \langle E_{\text{ini}}^S(x_1) | \text{Tr}_B [U^{\dagger SB} (\rho_{\text{can,fin}}^S \otimes \rho_{\text{can}}^B) U^{SB}] | E_{\text{ini}}^S(x_1) \rangle, \quad (47)$$

where $\rho_{\text{can,fin}}^S \otimes \rho_{\text{can}}^B$ is given by the right-hand side of equation (16). When the initially localized state expands into the total Hilbert space, the process would be absolutely irreversible and a positive entropy is produced during this process. The effect of absolute irreversibility is to lower the extractable work by $k_B T |\ln(1 - \lambda)|$.

We note that equation (43) without absolute irreversibility (i.e., $\lambda = 0$) can be derived by explicitly calculating the left-hand side of equation (43) as done in [13]. The calculation itself does not require the reference density matrix (16). However, the reason why the right-hand side of the Jarzynski equality (43) deviates from one can be clarified by introducing the reference probability distribution and the reference density matrix. Thus, the treatment of absolutely irreversible processes can be done in a physically transparent manner by introducing the reference probability distribution.

3. Nonequilibrium equalities with feedback control

3.1. Formulation of the problem

To realize a general measurement and a feedback protocol, we consider the following protocol which is basically the same as the one considered in [32] and schematically illustrated in figure 2.

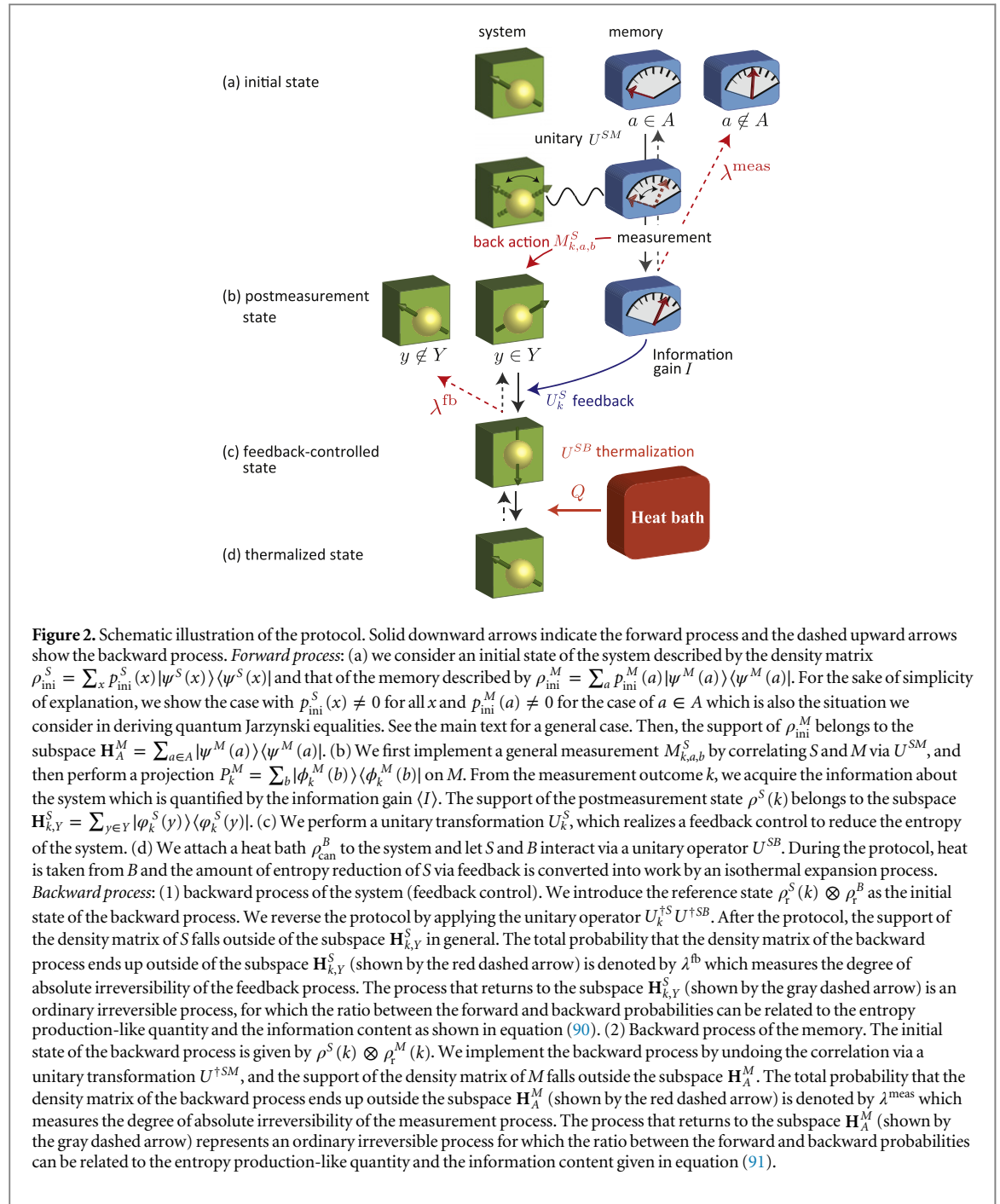
The total system consists of the system (S), the memory (M), the bath (B), and the interactions between them (SM and SB). The corresponding Hamiltonian is given by

$$H = H_k^S(t) + V^{SM}(t) + H^M + V^{SB}(t) + H^B, \quad (48)$$

where the interaction between the system and the heat bath is turned off until the isothermal expansion process (d) starts. The Hamiltonian of the system is controlled by the protocol that depends on the measurement outcome k after the measurement step (b) at time $t = t_{\text{meas}}$:

$$H_k^S(t) = \begin{cases} H^S(t), & 0 \leq t \leq t_{\text{meas}}; \\ H_k^S(t), & t_{\text{meas}} \leq t. \end{cases} \quad (49)$$

We denote the initial Hamiltonian of the system by $H_{\text{ini}}^S = H_k^S(t = 0)$.



(a) Let the initial state of the system and the memory be

$$\rho_{\text{ini}}^{SM} = \rho_{\text{ini}}^S \otimes \rho_{\text{ini}}^M. \quad (50)$$

(b) A general quantum measurement on the system is implemented by performing a unitary transformation $U^{SM} = \text{T exp} \left[-i \int_0^{t_{\text{meas}}} (H_k^S(t) + V^{SM}(t) + H^M) dt \right]$ between the system and the memory followed by a projection $P_k^M = \sum_b |\phi_k^M(b)\rangle \langle \phi_k^M(b)|$ on the memory. Here $P_k^M P_l^M = \delta_{k,l} P_k^M$ and $\{|\phi_k^M(b)\rangle\}_{k,b}$ is an orthonormal basis set of M . The postmeasurement state for the measurement outcome k is given by

$$\rho^S(k) = \frac{1}{p_k} P_k^M U^{SM} (\rho_{\text{ini}}^S \otimes \rho_{\text{ini}}^M) U^{+SM} P_k^M, \quad (51)$$

where $p_k = \text{Tr}[P_k^M U^{SM} (\rho_{\text{ini}}^S \otimes \rho_{\text{ini}}^M) U^{+SM} P_k^M]$ is the probability the outcome k being obtained. The reduced density matrix of the system $\rho^S(k) := \text{Tr}_M[\rho^{SM}(k)]$ is given by

$$\rho^S(k) = \sum_{a,b} \frac{M_{k,a,b}^S \rho^S M_{k,a,b}^{\dagger S}}{p_k}, \quad (52)$$

where

$$M_{k,a,b}^S = \sqrt{p_{\text{ini}}^M(a)} \langle \phi_k^M(b) | U^{SM} | \psi^M(a) \rangle \quad (53)$$

is the measurement operator satisfying the completeness relation

$$\sum_{k,a,b} M_{k,a,b}^{\dagger S} M_{k,a,b}^S = 1. \quad (54)$$

Here $|\psi^M(a)\rangle$ and $p_{\text{ini}}^M(a)$ in equation (53) are given by the spectral decomposition of the initial state of the memory: $\rho_{\text{ini}}^M = \sum_a p_{\text{ini}}^M(a) |\psi^M(a)\rangle \langle \psi^M(a)|$. Note that for the special case of $\rho_{\text{ini}}^M = |\psi^M(0)\rangle \langle \psi^M(0)|$ and $p_k^M = |\psi^M(k)\rangle \langle \psi^M(k)|$, the measurement is a pure measurement (which maps a pure state into a pure state)

$$M_k^S = \langle \psi^M(k) | U^{SM} | \psi^M(0) \rangle, \quad (55)$$

and the postmeasurement state is given by

$$\rho^S(k) = p_k^{-1} M_k^S \rho_{\text{ini}}^S M_k^{\dagger S}. \quad (56)$$

(c) We perform a unitary transformation U_k^S depending on the measurement outcome k . Here the unitary operator is given by $U_k^S = \text{T exp}[-i \int_{t_{\text{meas}}}^{t_{\text{fb}}} H_k^S(t) dt]$. We note that the above unitary operation associated with the measurement outcome is nothing but the feedback control. The density matrix of the system after the feedback control is given by

$$\rho_{\text{fb}}^S(k) = U_k^S \rho^S(k) U_k^{\dagger S}. \quad (57)$$

(d) Finally, we let the system and the heat bath interact with each other so that the reduced entropy of the system via feedback control is converted to heat. Here, we assume that the initial state of the heat bath is given by the canonical distribution, i.e., $\rho_{\text{can}}^B = \exp[-\beta(H^B - F^B)]$. The final state is given by

$$\rho_{\text{fin}}^{SB}(k) = U_k^{SB} (\rho_{\text{fb}}^S(k) \otimes \rho_{\text{can}}^B) U_k^{\dagger SB}, \quad (58)$$

where the interaction between S and B is described by the unitary operator

$U_k^{SB} = \text{T exp}[-i \int_{t_{\text{fb}}}^{t_{\text{fin}}} (H_k^S(t) + V^{SB}(t) + H^B) dt]$, which may, in general, depend on k .

Now we introduce reference states for each subsystem and define entropy production-like quantities which measure the amount of entropy of SB (M) that is reduced (or produced) due to the feedback control (or measurement). The reference states of each subsystem are given by

$$\rho_r^S(k) = \sum_z p_r^S(z|k) |\phi_k^S(z)\rangle \langle \phi_k^S(z)|, \quad (59)$$

$$\rho_r^M(k) = \sum_b p_r^M(b|k) |\phi_k^M(b)\rangle \langle \phi_k^M(b)|, \quad (60)$$

$$\rho_r^B = \sum_j p_{\text{can}}^B(j) |\psi^B(j)\rangle \langle \psi^B(j)|, \quad (61)$$

where $p_{\text{can}}^B(j) = e^{-\beta(E^B(j) - F^B)}$ is the canonical distribution, and $E^B(j)$ and $|\psi^B(j)\rangle$ are the eigenenergy and the energy eigenstate of the heat bath, respectively.

We define the following quantity that measures the amount of entropy reduction of SB due to feedback control:

$$\langle \sigma^{SB} \rangle = -S(\rho_{\text{ini}}^S \otimes \rho_{\text{can}}^B) - \sum_k p_k \text{Tr}[\rho_{\text{fin}}^{SB}(k) \ln(\rho_r^S(k) \otimes \rho_r^B)] \quad (62)$$

$$= -S(\rho_{\text{ini}}^S) - \sum_k p_k \text{Tr}[\rho_{\text{fin}}^S(k) \ln \rho_r^S(k)] - \beta \langle Q \rangle, \quad (63)$$

where

$$\langle Q \rangle = \text{Tr}[H^B \rho_{\text{can}}^B] - \sum_k p_k \text{Tr}[H^B \rho_{\text{fin}}^B(k)] \quad (64)$$

is the energy change of the heat bath which we identify as heat transferred from B to S . Note that if we choose the reference state as the final density matrix of S , i.e., $\rho_r^S(k) = \rho_{\text{fin}}^S(k)$, equation (63) is nothing but the total entropy change of SB due to feedback control:

$$\langle \sigma^{SB} \rangle = \Delta S^S - \beta \langle Q \rangle, \quad (65)$$

where

$$\Delta S^S = \sum_k p_k S(\rho_{\text{fin}}^S(k)) - S(\rho_{\text{ini}}^S) \quad (66)$$

is a change in the von Neumann entropy of the system during the entire protocol.

We also define the following quantity which measures the amount of entropy produced in M due to measurement:

$$\langle \sigma^M \rangle = -S(\rho_{\text{ini}}^M) - \text{Tr}[\rho_{\text{fin}}^M \ln \rho_{\text{r}}^M], \quad (67)$$

where $\rho_{\text{fin}}^M = \sum_k p_k \rho^M(k)$ is the final density matrix of M and $\rho_{\text{r}}^M := \sum_k p_k \rho_{\text{r}}^M(k)$. If we choose the reference state as the canonical distribution, equations (63) and (67) are related to work and the free-energy difference, respectively, as shown in the next section.

The entropy production-like quantities (63) and (67) contain not only the effect of dissipated entropy due to irreversibility of the process but also the effect of entropy change due to information processing (measurement and feedback control), and they can take either positive or negative values depending on the process. The effect of the information exchange between the system and the memory can be expressed by the information gain (quantum-classical mutual information) of the system S [47, 51, 52]:

$$\langle I \rangle = S(\rho_{\text{ini}}^S) - \sum_k p_k S(\rho^S(k)), \quad (68)$$

which is the amount of entropy that is reduced from the system due to the measurement. The information gain is bounded from above by the Shannon entropy $H = -\sum_k p_k \ln p_k$, i.e., $\langle I \rangle \leq H$, where the equality holds if and only if the measurement is given by a projective measurement using the diagonal basis of ρ_{ini}^S . Moreover, the information gain is nonnegative for any premeasurement state if the measurement is a pure measurement (55) as discussed in [52].

Extracting the information gain from entropy production-like quantities (63) and (67), we obtain the measures of irreversibility during measurement and feedback processes. For the feedback process, we have

$$\begin{aligned} \langle \sigma^{SB} \rangle + \langle I \rangle &= -S(\rho_{\text{can}}^B) - \sum_k p_k S(\rho^S(k)) - \sum_k p_k \text{Tr}[\rho_{\text{fin}}^{SB}(k) \ln(\rho_{\text{r}}^S(k) \otimes \rho_{\text{r}}^B)] \\ &= -\sum_k p_k \left\{ S(\rho^S(k) \otimes \rho_{\text{can}}^B) + \text{Tr}[(\rho^S(k) \otimes \rho_{\text{can}}^B) \ln \tilde{\rho}_{\text{r}}^{SB}(k)] \right\} \\ &= \sum_k p_k S(\rho^S(k) \otimes \rho_{\text{can}}^B \parallel \tilde{\rho}_{\text{r}}^{SB}(k)) \geq 0, \end{aligned} \quad (69)$$

where $\tilde{\rho}_{\text{r}}^{SB}(k) = U_k^{\dagger S} U_k^{\dagger SB} (\rho_{\text{r}}^S(k) \otimes \rho_{\text{r}}^B) U_k^{SB} U_k^S$ is the final density matrix of the backward process by reversing the isothermal expansion and feedback control protocols. Note that the feedback protocol of the system (and the heat bath) is reversible if and only if $\rho^S(k) \otimes \rho_{\text{can}}^B = \tilde{\rho}_{\text{r}}^{SB}(k)$ [57], which is the equality condition of the last inequality (69), that is $\langle \sigma^{SB} \rangle + \langle I \rangle = 0$.

Similarly, for the measurement process, we have

$$\begin{aligned} \langle \sigma^M \rangle - \langle I \rangle &= -S(\rho_{\text{ini}}^{SM}) - \text{Tr}[\rho_{\text{fin}}^M \ln \rho_{\text{r}}^M] + \sum_k p_k S(\rho^S(k)) \\ &= -S(U^{SM} \rho_{\text{ini}}^{SM} U^{\dagger SM}) - \text{Tr} \left[\rho_{\text{meas}}^{SM} \ln \left(\sum_k p_k \rho^S(k) \otimes \rho_{\text{r}}^M(k) \right) \right] \\ &= \Delta S_{\text{meas}}^{SM} + S \left(\rho_{\text{meas}}^{SM} \parallel \sum_k p_k \rho^S(k) \otimes \rho_{\text{r}}^M(k) \right) \geq 0, \end{aligned} \quad (70)$$

where

$$\rho_{\text{meas}}^{SM} = \sum_k p_k \rho^{SM}(k) \quad (71)$$

is the density operator of the postmeasurement state averaged over measurement outcomes, and

$$\Delta S_{\text{meas}}^{SM} = S(\rho_{\text{meas}}^{SM}) - S(U^{SM} \rho_{\text{ini}}^{SM} U^{\dagger SM}) \geq 0 \quad (72)$$

is a change in the von Neumann entropy due to projection P_k^M and the inequality results from the fact that the von Neumann entropy does not decrease under projection measurements. The nonnegativity in equation (70) shows the irreversibility of the measurement process.

Next, let us consider the following spectral decompositions of the initial states of the system, the heat bath, and the memory

$$\rho_{\text{ini}}^S = \sum_x p_{\text{ini}}^S(x) |\psi^S(x)\rangle \langle \psi^S(x)|, \quad (73)$$

$$\rho_{\text{ini}}^B = \sum_i p_{\text{can}}^B(h) |\psi^B(h)\rangle \langle \psi^B(h)|, \quad (74)$$

$$\rho_{\text{ini}}^M = \sum_a p_{\text{ini}}^M(a) |\psi^M(a)\rangle \langle \psi^M(a)|. \quad (75)$$

Let us also decompose the postmeasurement state of the system as follows:

$$\rho^S(k) = \sum_y p^S(y|k) |\varphi_k^S(y)\rangle \langle \varphi_k^S(y)|. \quad (76)$$

Using the above decompositions, we calculate equations (69) and (70) and define an unaveraged form of equations (63), (67) and (68), along a line similar to what we did in deriving the unaveraged form of the entropy production in equation (25). From equation (62), we obtain

$$\begin{aligned} \langle \sigma^{SB} \rangle &= \sum_x p_{\text{ini}}^S(x) \ln p_{\text{ini}}^S(x) + \sum_h p_{\text{can}}^B(h) \ln p_{\text{can}}^B(h) \\ &\quad - \sum_{k,j,z} p_k \langle \phi_k^S(z) | \otimes \langle \psi^B(j) | \rho_{\text{fin}}^{SB}(k) | \phi_k^S(z) \rangle \otimes | \psi^B(j) \rangle \ln (p_r^S(z|k) p_{\text{can}}^B(j)) \\ &= \sum_{x,a,h,k,y,b,j,z} p(x, a, h, k, y, b, j, z) \ln \frac{p_{\text{ini}}^S(x) p_{\text{can}}^B(h)}{p_r^S(z|k) p_{\text{can}}^B(j)}, \end{aligned} \quad (77)$$

where we introduce the forward probability distribution corresponding to the forward process of the total system:

$$p(x, a, h, k, y, b, j, z) = p_{\text{ini}}^S(x) p_{\text{ini}}^M(a) p_{\text{can}}^B(h) p(k, y, b|x, a) p(z, j|k, y, h), \quad (78)$$

where

$$\begin{aligned} p(k, y, b|x, a) &= \left| \langle \varphi_k^S(y) | \otimes \langle \phi_k^M(b) | U^{SM} | \psi^S(x) \rangle \otimes | \psi^M(a) \rangle \right|^2 \\ &= \frac{1}{p_{\text{ini}}^M(a)} \left| \langle \varphi_k^S(y) | M_{k,a,b}^S | \psi^S(x) \rangle \right|^2 \end{aligned} \quad (79)$$

is the transition probability from the state labeled by x, a to the state labeled by k, y, b during the measurement process, and

$$p(z, j|k, y, h) = \left| \langle \phi_k^S(z) | \otimes \langle \psi^B(j) | U_k^{SB} U_k^S | \varphi_k^S(y) \rangle \otimes | \psi^B(h) \rangle \right|^2 \quad (80)$$

is the transition probability from the state labeled by k, y, h to the state labeled by z, j during the feedback and the isothermal expansion protocol. Note that in deriving the last line of equation (77), we used the relation

$$\begin{aligned} &\langle \phi_k^S(z) | \otimes \langle \psi^B(j) | \rho_{\text{fin}}^{SB}(k) | \phi_k^S(z) \rangle \otimes | \psi^B(j) \rangle \\ &= \sum_{h,y} p(z, j|k, y, h) p^S(y|k) p_{\text{can}}^B(h) \\ &= \sum_{x,a,h,y,b} p_k^{-1} p_{\text{ini}}^S(x) p_{\text{ini}}^M(a) p_{\text{can}}^B(h) p(k, y, b|x, a) p(z, j|k, y, h). \end{aligned} \quad (81)$$

We also follow the same procedure for the memory using equation (67) and obtain

$$\begin{aligned} \langle \sigma^M \rangle &= \sum_a p_{\text{ini}}^M(a) \ln p_{\text{ini}}^M(a) - \sum_{k,b} p_k \langle \phi_k^M(b) | \rho^M(k) | \phi_k^M(b) \rangle \ln (p_k p_r^M(b|k)) \\ &= \sum_{x,a,k,y,b} p^{\text{meas}}(x, a, k, y, b) \ln \frac{p_{\text{ini}}^M(a)}{p_k p_r^M(b|k)}, \end{aligned} \quad (82)$$

where the forward probability of the measurement process is defined as

$$p^{\text{meas}}(x, a, k, y, b) = p_{\text{ini}}^S(x) p_{\text{ini}}^M(a) p(k, y, b|x, a). \quad (83)$$

From equations (68), (77) and (82), we define unaveraged entropy production-like quantities and the corresponding information content as follows:

$$\sigma^{SB}(x, h, k, j, z) = \ln[p_{\text{ini}}^S(x)p_{\text{can}}^B(h)] - \ln[p_r^S(z|k)p_{\text{can}}^B(j)] \quad (84)$$

$$= \ln p_{\text{ini}}^S(x) - \ln p_r^S(z|k) - \beta Q(h, j), \quad (85)$$

$$\sigma^M(a, k, b) = \ln p_{\text{ini}}^M(a) - \ln[p_k p_r^M(b|k)], \quad (86)$$

$$I(x, k, y) = \ln p^S(y|k) - \ln p_{\text{ini}}^S(x), \quad (87)$$

where $Q(h, j) = E^B(h) - E^B(j)$ is the heat transferred from the heat bath to the system. Since the entropy production relates the forward and reference probabilities as in equation (32), we have similar relations for the combinations $\sigma^{SB} + I$ and $\sigma^M - I$. First by decomposing equation (69), we have

$$\langle \sigma^{SB} \rangle + \langle I \rangle = \sum_{h,k,y,j,z} p^{\text{fb}}(h, k, y, j, z) \ln \frac{p^S(y|k)p_{\text{can}}^B(h)}{p_r^S(z|k)p_{\text{can}}^B(j)}, \quad (88)$$

where

$$p^{\text{fb}}(h, k, y, j, z) = p_{\text{can}}^B(h)p^S(y|k)p_k p(z, j|k, y, h) \quad (89)$$

is the forward probability of the feedback control process. Then, $\sigma^{SB} + I$ and $\sigma^M - I$ are expressed in terms of the logarithm of the ratio between the forward and reference probabilities as

$$\sigma^{SB}(x, h, k, j, z) + I(x, k, y) = \ln \frac{p^{\text{fb}}(h, k, y, j, z)}{p_r^{\text{fb}}(h, k, y, j, z)}, \quad (90)$$

$$\sigma^M(a, k, b) - I(x, k, y) = \ln \frac{p_r^{\text{meas}}(x, a, k, y, b)}{p_r^{\text{meas}}(x, a, k, y, b)}, \quad (91)$$

where

$$p_r^{\text{fb}}(h, k, y, j, z) = p_k p_r^S(z|k)p_{\text{can}}^B(j)p(z, j|k, y, h) \quad (92)$$

and

$$p_r^{\text{meas}}(x, a, k, y, b) = p_k p^S(y|k)p_r^M(b|k)p(k, y, b|x, a) \quad (93)$$

are the reference probabilities of the feedback and measurement processes, respectively.

Note that equation (92) gives the probability of the system and the heat bath returning to the postmeasurement state of the system and the initial state of the heat bath $|\varphi_k^S(y)\rangle \otimes |\psi^B(h)\rangle$ when we start from the initial state of the backward process $\rho_r^S(z|k) \otimes \rho_{\text{can}}^B$ and do the reverse of the isothermal expansion and feedback control $U_k^{\dagger S} U_k^{\dagger SB}$, as shown in the gray dashed upward arrow in figure 2. Also, equation (93) gives the probability of the system and the memory returning to the initial state $|\psi^S(x)\rangle \otimes |\psi^M(a)\rangle$ when we start from the initial state of the backward process $\sum_k p_k \rho^S(k) \otimes \rho_r^M(k)$ and let the system and the memory undo the correlation by applying a unitary operation $U^{\dagger SM}$, as shown by the gray dashed upward arrow in figure 2. We use the definitions of entropy production-like quantities (85) and (86) and the information content (87) to derive quantum fluctuation theorems for both the feedback-controlled system and the measurement device.

3.2. Quantum fluctuation theorems with feedback control

We derive quantum fluctuation theorems for both the feedback-controlled system and the measurement device from the fact that the sum of the reference probabilities is unity for both the feedback control process (92) and the measurement process (93):

$$1 = \sum_{h,k,y,j,z} p_r^{\text{fb}}(h, k, y, j, z), \quad (94)$$

$$1 = \sum_{x,a,y,k,b} p_r^{\text{meas}}(x, a, k, y, b). \quad (95)$$

As in equation (34), we decompose equations (94) and (95) into two parts; one is the part where we can take the ratio between the forward and reference probabilities, and the other is the part where the corresponding forward probability vanishes. See dashed upward arrows in figure 2.

We introduce a set of labels corresponding to the non-vanishing probability distributions as follows: we introduce Y as a set of labels y satisfying $p(y|k) \neq 0$, and A as a set of labels (x, a) satisfying both $p_{\text{ini}}^S(x) \neq 0$ and $p_0^M(a) \neq 0$. Then, the support of the postmeasurement state of the system belongs to the subspace $\mathbf{H}_{k,Y}^S$, which is spanned by $\{|\varphi_k^S(y)\rangle\}_{y \in Y}$, and the support of the initial state of the system and the memory belongs to the subspace \mathbf{H}_A^{SM} , which is spanned by $\{|\psi^S(x)\rangle \otimes |\psi^M(a)\rangle\}_{(x,a) \in A}$.

Using the above notations, we decompose the reference states and derive quantum fluctuation theorems as follows: for the system and the heat bath, we have

$$\begin{aligned}
 1 &= \sum_{h,k,y \notin Y,j,z} p_r^{\text{fb}}(h, k, y, j, z) + \sum_{h,k,y \in Y,j,z} p_r^{\text{fb}}(h, k, y, j, z) \\
 &= \lambda^{\text{fb}} + \sum_{h,k,y \in Y,j,z} p^{\text{fb}}(h, k, y, j, z) \frac{p_r^{\text{fb}}(h, k, y, j, z)}{p^{\text{fb}}(h, k, y, j, z)} \\
 &= \lambda^{\text{fb}} + \sum_{x,a,h,k,y \in Y,b,j,z} p(x, a, h, k, y, b, j, z) e^{-\sigma^{\text{SB}}(x,h,k,j,z) - I(x,k,y)} \\
 &= \lambda^{\text{fb}} + \left\langle e^{-\sigma^{\text{SB}} - I} \right\rangle,
 \end{aligned} \tag{96}$$

where we used equation (90) in deriving the third equality. Here

$$\lambda^{\text{fb}} = \sum_{h,k,y \notin Y,j,z} p_r^{\text{fb}}(h, k, y, j, z) \tag{97}$$

is the sum of the reference probabilities, where the density matrix of the backward process ends up outside of the subspace $\mathbf{H}_{k,Y}^S$, and the overlap with the postmeasurement state of the forward process is zero.

For the memory, the decomposition of the reference probability leads to

$$\begin{aligned}
 1 &= \sum_{(x,a) \notin A,k,y,b} p_r^{\text{meas}}(x, a, k, y, b) + \sum_{(x,a) \in A,k,y,b} p_r^{\text{meas}}(x, a, k, y, b) \\
 &= \lambda^{\text{meas}} + \sum_{(x,a) \in A,h,k,y,b,j,z} p(x, a, h, k, y, b, j, z) \frac{p_r^{\text{meas}}(x, a, k, y, b)}{p^{\text{meas}}(x, a, k, y, b)} \\
 &= \lambda^{\text{meas}} + \sum_{(x,a) \in A,h,k,y,b,j,z} p(x, a, h, k, y, b, j, z) e^{-\sigma^M(a,k,b) + I(x,k,y)} \\
 &= \lambda^{\text{meas}} + \left\langle e^{-\sigma^M + I} \right\rangle,
 \end{aligned} \tag{98}$$

where we used equation (91) in deriving the third equality. Here

$$\lambda^{\text{meas}} = \sum_{(x,a) \notin A,k,y,b} p_r^{\text{meas}}(x, a, k, y, b) \tag{99}$$

is the sum of the reference probabilities such that the density matrix of the backward process ends up outside of the subspace \mathbf{H}_A^{SM} and the overlap with the initial state of the forward process vanishes.

Rewriting equations (96) and (98), we obtain the quantum fluctuation theorems for the system and the memory:

$$\left\langle e^{-\sigma^{\text{SB}} - I} \right\rangle = 1 - \lambda^{\text{fb}}, \tag{100}$$

$$\left\langle e^{-\sigma^M + I} \right\rangle = 1 - \lambda^{\text{meas}}. \tag{101}$$

Using the Jensen's inequality, we can reproduce second law-like inequalities by using equations (100) and (101):

$$\langle \sigma^{\text{SB}} \rangle \geq -\langle I \rangle - \ln(1 - \lambda^{\text{fb}}), \tag{102}$$

$$\langle \sigma^M \rangle \geq \langle I \rangle - \ln(1 - \lambda^{\text{meas}}), \tag{103}$$

where the presence of absolute irreversibility (i.e., nonzero λ) imposes stronger lower bounds on $\langle \sigma^{\text{SB}} \rangle$ and $\langle \sigma^M \rangle$ compared with the previous results (69) and (70) given in [49]. Since λ^{fb} gives the total probability of the density matrix of the backward process ending up outside of the subspace $\mathbf{H}_{k,Y}^S$, it measures the degree of absolute irreversibility of the feedback protocol.

If the measurement on the system is given by projective measurements $|k\rangle\langle k|_S$, the situation becomes simple. In this case, λ^{fb} takes the following form:

$$\begin{aligned}
 \lambda^{\text{fb}} &= \sum_{h,k \neq y,j,z} p_k p_r^S(z|k) \rho_{\text{can}}^B(j) \left| \langle y|_S \otimes \langle \psi^B(h) | U_k^{\dagger S} U_k^{\dagger \text{SB}} | \phi_k^S(z) \rangle \otimes | \psi^B(j) \rangle \right|^2 \\
 &= \sum_{k \neq y} p_k \left\langle y \left| \text{Tr}_B \left[U_k^{\dagger S} U_k^{\dagger \text{SB}} \left(\rho_r^S(k) \otimes \rho_{\text{can}}^B \right) U_k^{\text{SB}} U_k^S \right] \right| y \right\rangle_S,
 \end{aligned} \tag{104}$$

which is the sum of the probability of the backward protocol for each measurement outcome k that does not end in the state $|k\rangle_S$. If the unitary operator brings the postmeasurement state $|k\rangle_S$ into the reference state $\rho^S(k)$ for

all k , the feedback (and isothermal expansion) process is reversible and λ^{fb} vanishes; otherwise the irreversibility of the process reduces the efficiency of the feedback gain. Note that equation (100) holds even for projective measurements on the system, whereas the previous results in [32] are inapplicable, since we take into account the effect of absolute irreversibility. Although the obtained information is given by the Shannon entropy and is maximal for projective measurements, feedback protocol tends to be absolutely irreversible since the postmeasurement state, which is given by a pure state $|k\rangle$, is sharply localized in the Hilbert space. Similarly, λ^{meas} measures the absolute irreversibility of the measurement process since it is nonzero when the density matrix of the backward protocol ends up outside of the subspace \mathbf{H}_A^{SM} .

Now let us compare the obtained equalities (100) and (101) with the quantum fluctuation theorems of the total system by using the total entropy production $\sigma_{\text{tot}}(x, h, k, j, a, b) = \sigma^{\text{SB}}(x, h, k, j, a) + \sigma^{\text{M}}(a, k, b)$. The total entropy production can be written in the form

$$\begin{aligned}\sigma_{\text{tot}}(x, h, k, j, a, b) &= \ln \frac{p_{\text{ini}}^{\text{S}}(x)p_{\text{can}}^{\text{B}}(h)p_{\text{ini}}^{\text{M}}(a)}{p_{\text{r}}^{\text{S}}(z|k)p_{\text{can}}^{\text{B}}(j)p_k p_{\text{r}}^{\text{M}}(b|k)} \\ &= \ln \frac{p(x, a, h, k, b, j, z)}{p_{\text{r}}(x, a, h, k, b, j, z)},\end{aligned}\quad (105)$$

where we used the total probability distribution of the forward process (78), and

$$p_{\text{r}}(x, a, h, k, b, j, z) = p_{\text{r}}^{\text{S}}(z|k)p_{\text{can}}^{\text{B}}(j)p_k p_{\text{r}}^{\text{M}}(b|k)p(z, j|k, \gamma, h)p(k, \gamma, b|x, a) \quad (106)$$

is the total probability distribution of the backward process. Since the sum of the reference probability (106) is unity, we can derive the quantum fluctuation theorem for the total system:

$$\begin{aligned}1 &= \sum_{(x,a) \notin A,h,k,b,j,z} p_{\text{r}}(x, a, h, k, b, j, z) + \sum_{(x,a) \in A,h,k,b,j,z} p_{\text{r}}(x, a, h, k, b, j, z) \\ &= \lambda^{\text{tot}} + \sum_{(x,a) \in A,h,k,b,j,z} p(x, a, h, k, b, j, z) \ln \frac{p_{\text{r}}(x, a, h, k, b, j, z)}{p(x, a, h, k, b, j, z)} \\ &= \lambda^{\text{tot}} + \langle e^{-\sigma_{\text{tot}}} \rangle.\end{aligned}\quad (107)$$

Since the obtained fluctuation theorem is applicable to the total system, the effect of information exchange between S and M is canceled, and the information content does not appear in equation (107). Moreover, λ^{tot} measures absolute irreversibility of the combined process of the measurement and feedback control, whereas from equations (100) and (101) we can separately obtain the information about absolute irreversibility in measurement and feedback.

3.3. Quantum Jarzynski equalities with feedback control

In this subsection, we derive the quantum Jarzynski equality for the feedback-controlled system by assuming that the initial and reference states are given by canonical distributions

$$\rho_{\text{ini}}^{\text{S}} = e^{-\beta(H_{\text{ini}}^{\text{S}} - F_{\text{ini}}^{\text{S}})}, \quad (108)$$

and

$$\rho_{\text{r}}^{\text{S}}(k) = e^{-\beta(H_{\text{fin}}^{\text{S}}(k) - F_k^{\text{S}})}, \quad (109)$$

where $H_{\text{ini}}^{\text{S}}$ and $H_{\text{fin}}^{\text{S}}(k)$ are the initial and final Hamiltonians of the system. Then, the orthogonal bases $\{|\psi^{\text{S}}(x)\rangle\}$ and $\{|\phi_k^{\text{S}}(z)\rangle\}$ are given by the set of energy eigenfunctions: $H_{\text{ini}}^{\text{S}}|\psi^{\text{S}}(x)\rangle = E_{\text{ini}}^{\text{S}}(x)|\psi^{\text{S}}(x)\rangle$ and $H_{\text{fin}}^{\text{S}}(k)|\phi_k^{\text{S}}(z)\rangle = E_{\text{fin},k}^{\text{S}}(z)|\phi_k^{\text{S}}(z)\rangle$. Now σ^{SB} is related to the work done by the system as follows:

$$\sigma^{\text{SB}}(x, h, k, j, z) = -\beta[W^{\text{S}}(x, h, k, j, z) + \Delta f^{\text{S}}(k)], \quad (110)$$

where

$$W^{\text{S}}(x, k, z) = E_{\text{ini}}^{\text{S}}(x) - E_{\text{fin},k}^{\text{S}}(z) + Q(h, j) \quad (111)$$

is the work done by the system, and $\Delta f^{\text{S}}(k) = F_k^{\text{S}} - F_{\text{ini}}^{\text{S}}$ is the free-energy difference.

We now derive the following quantum Jarzynski equality for a feedback-controlled system by using equation (100):

$$\left\langle e^{\beta(W^{\text{S}} + \Delta f^{\text{S}}) - I} \right\rangle = 1 - \lambda^{\text{fb}}. \quad (112)$$

Using the Jensen's inequality, equation (112) reproduces the generalized second law under feedback control:

$$\langle W^S \rangle \leq -\langle \Delta f^S \rangle + k_B T \langle I \rangle + k_B T \ln(1 - \lambda^{\text{fb}}), \quad (113)$$

where

$$\langle W^S \rangle = \text{Tr}[\rho_{\text{ini}}^S H_{\text{ini}}^S] - \sum_k p_k \text{Tr}[\rho_{\text{fin}}^S(k) H_{\text{fin}}^S(k)] + \langle Q \rangle \quad (114)$$

is the averaged work done by the system. Imperfect feedback control leads to nonzero λ^{fb} , which lowers the extractable work from the system as shown in equation (113).

Next, we derive the quantum Jarzynski equality for the memory that acquires the measurement results by assuming that the initial and reference states are given by the canonical distributions

$$\rho_{\text{ini}}^M = e^{-\beta(H_0^M - F_0^M)}, \quad (115)$$

$$\rho_{\text{r}}^M(k) = e^{-\beta(H_k^M - F_k^M)}. \quad (116)$$

We assume that the initial state of the memory (115) is given by the local equilibrium state defined as the canonical distribution using a local Hamiltonian H_0^M , where the Hamiltonian of the memory is decomposed into $H^M = \bigoplus_k H_k^M$ [48]. Here, the spectral decomposition of each local Hamiltonian is given by $H_k^M = \sum_b E_k^M(b) |\phi_k^M(b)\rangle \langle \phi_k^M(b)|$. For convenience, let us relabel a as $a = (a_1, a_2)$ so that $|\psi^M(a)\rangle = |\phi_{a_1}^M(a_2)\rangle$. Then $p_{\text{ini}}^M(a) \neq 0$ if $a = (0, a_2)$ and $p_{\text{ini}}^M(a) = 0$ otherwise for the initial state defined in equation (115). Now σ^M is related to the work done on the memory as follows:

$$\sigma^M(a, k, b) = -\beta(W^M(a, k, b) + \Delta f^M(k) + H(k)), \quad (117)$$

where

$$W^M(a, k, b) = E_0^M(a_2) - E_k^M(b) \quad (118)$$

is the work done by the memory, $\Delta f^M(k) = F_k^M - F_0^M$ is the free-energy difference, and $H(k) = -\ln p_k$ is the (unaveraged) Shannon entropy. We can derive the following quantum Jarzynski equality for the memory by using equation (101):

$$\left\langle e^{\beta(W^M + \Delta f^M) - H + I} \right\rangle = 1 - \lambda^{\text{meas}}. \quad (119)$$

Using the Jensen's inequality, equation (119) reproduces the generalized second law for the memory:

$$\langle W^M \rangle \leq -\langle \Delta f^M \rangle - k_B T (\langle I \rangle - \langle H \rangle) + k_B T \ln(1 - \lambda^{\text{meas}}), \quad (120)$$

where $\langle W^M \rangle = \text{Tr}[\rho_0^M H_0^M] - \sum_k p_k \text{Tr}[\rho_k^M H_k^M]$ is the averaged work done by the memory and $\langle H \rangle = -\sum_k p_k \ln p_k$ is the Shannon entropy. Note that $\langle W^M \rangle$ usually takes a negative value since we need to input energy to the memory to perform the measurement. A nonzero λ^{meas} increases the work cost of the measurement due to absolute irreversibility as shown in equation (120). Using the setup of our Hamiltonian of the memory in this section, λ^{meas} can be expressed as

$$\lambda^{\text{meas}} = \sum_{x, a_1 \neq 0, a_2, k, y, b} p_k p^S(y|k) p_{\text{can}}^M(b|k) \left| \langle \psi^S(x) | \otimes \langle \phi_{a_1}^M(a_2) | U^{\dagger SM} | \varphi_k^S(y) \rangle \otimes | \phi_k^M(b) \rangle \right|^2, \quad (121)$$

where $p_{\text{can}}^M(b|k) = \exp[-\beta(E_k^M(b) - F_k^M)]$ is the canonical distribution corresponding to the initial state of the backward process, and $A = \{(x, a) | a = (0, a_2)\}$ since $p_{\text{can}}^S(x) \neq 0$ for all x in this setup. From equation (121), we note that λ^{meas} is the total probability that the backward process ends in the subspace $\{|\phi_{a_1}^M(a_2)\rangle\}_{a_1 \neq 0, a_2}$, which was not occupied by the initial local equilibrium state (115). Note that the projection on the memory destroys the coherence between the system and the memory, which occurs only in the quantum regime due to the measurement back action.

3.4. Quantum fluctuation theorems for feedback-controlled systems and unavailable information

In this section, we consider the effect of absolute irreversibility during the feedback process in more detail. Without absolute irreversibility, the extra work beyond the conventional second law of thermodynamics that can be extracted from the system is bounded from above by $k_B T$ times the obtained information $\langle I \rangle$. However, if the feedback process is absolutely irreversible, we cannot fully utilize the information to extract work. We introduce the amount of information that is unavailable for use in extracting work for a given feedback control protocol as

$$I_u(k) = -\ln(1 - \lambda^{\text{fb}}(k)). \quad (122)$$

This quantity was originally introduced in [45] for classical systems. Here, we introduce the total reference probability of the system not going back to the postmeasurement state conditioned on the measurement outcome k :

$$\lambda^{\text{fb}}(k) = \frac{1}{p_k} \sum_{h,y \notin Y_{j,z}} p_r^{\text{fb}}(h, k, y, j, z). \quad (123)$$

Then we start from the following relation:

$$p_k = \sum_{h,y \notin Y_{j,z}} p_r^{\text{fb}}(h, k, y, j, z) + \sum_{h,y \in Y_{j,z}} p_r^{\text{fb}}(h, k, y, j, z). \quad (124)$$

From equations (122) and (123), we find that equation (124) takes the form

$$\begin{aligned} \sum_{h,y \in Y_{j,z}} p_r^{\text{fb}}(h, k, y, j, z) &= p_k - p_k \lambda^{\text{fb}}(k) \\ &= p_k e^{-I_u(k)}. \end{aligned} \quad (125)$$

Multiplying both sides of equation (125) by $e^{I_u(k)}$, summing over k and using equation (90), we obtain

$$\left\langle e^{-\sigma^{\text{SB}} - (I - I_u)} \right\rangle = 1. \quad (126)$$

Using Jensen's inequality, we obtain the inequality for σ^{SB} in the presence of unavailable information:

$$\langle \sigma^{\text{SB}} \rangle \geq -(\langle I \rangle - \langle I_u \rangle). \quad (127)$$

If we use the same assumptions (108) and (109) in deriving the quantum Jarzynski equality, we obtain

$$\left\langle e^{\beta(W^S + \Delta f^S) - (I - I_u)} \right\rangle = 1, \quad (128)$$

and hence

$$\langle W^S \rangle \leq -\langle \Delta f^S \rangle + k_B T (\langle I \rangle - \langle I_u \rangle). \quad (129)$$

The obtained inequalities (127) and (129) give bounds on the entropy reduction of SB and extractable work from the system, where they take into account the inefficiency of the feedback control by subtracting the unavailable information $\langle I_u \rangle$ from the obtained information $\langle I \rangle$. From the convexity, the unavailable information is bounded from above by

$$-I_u = \sum_k p_k \ln(1 - \lambda^{\text{fb}}(k)) \leq \ln\left(1 - \sum_k p_k \lambda^{\text{fb}}(k)\right) = \ln(1 - \lambda^{\text{fb}}), \quad (130)$$

so that inequality (127) gives a tighter bound compared with inequality (102).

4. Examples

4.1. Quantum piston

In this subsection, we consider a free expansion of a gas by using a model of the quantum piston discussed in [43, 44], and apply the quantum Jarzynski equality (43) to this model.

We consider a quantum particle with mass M trapped inside a one-dimensional box whose length is given by L_t at time t . We start with $L_0 = A$ and pull the piston with a constant speed v . At the final time, the length of the box is given by $L_{t_f} = B > A$, meaning that $L_t = vt + A$ with $t_f = (B - A)/v$. See figure 3. The instantaneous energy eigenstate is denoted by $|m^L\rangle$, and its wave function is given by

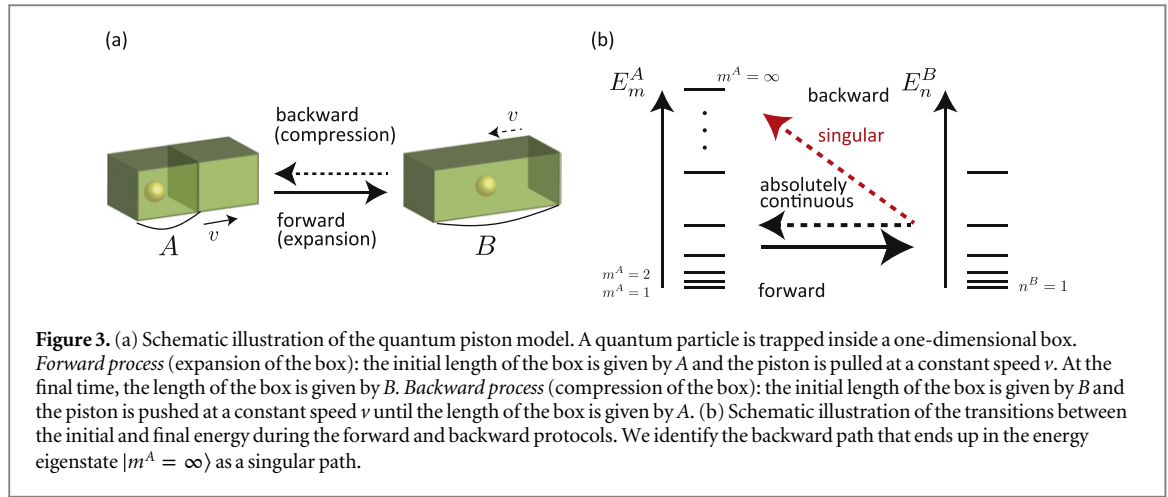
$$\phi_m(x; L) = \sqrt{\frac{2}{L}} \sin\left(\frac{m\pi x}{L}\right), \quad (131)$$

and the corresponding eigenenergy is given by

$$E_m^L = \frac{m^2 \pi^2 \hbar^2}{2ML^2}. \quad (132)$$

We consider a process starting with a canonical distribution

$$\rho_{\text{ini}} = \sum_m p_{\text{ini}}(m^A) |m^A\rangle \langle m^A|, \quad (133)$$



where

$$p_{\text{ini}}(m^A) = \frac{1}{Z_A} e^{-\beta E_m^A}, \quad (134)$$

with $Z_A = \sum_m \exp(-\beta E_m^A)$ being the partition function. The conditional probability distribution of obtaining the final state $|n^B\rangle$ for a given initial state $|m^A\rangle$ is given by

$$p(n^B|m^A) = \left| \langle n^B | U | m^A \rangle \right|^2, \quad (135)$$

where U is the unitary operator describing the time evolution of the system. An explicit form of $p(n^B|m^A)$ is given in [43] as

$$p(n^B|m^A) = \left| \sum_{l=1}^{\infty} \frac{2}{A} \int_0^A e^{-\frac{iMvx^2}{2A\hbar}} \sin\left(\frac{l\pi x}{A}\right) \sin\left(\frac{m\pi x}{A}\right) dx e^{-\frac{i\pi^2 l^2 \hbar (B-A)}{2ABMv}} \times \frac{2}{B} \int_0^B e^{\frac{iMvy^2}{2B\hbar}} \sin\left(\frac{l\pi y}{B}\right) \sin\left(\frac{n\pi y}{B}\right) dy \right|^2. \quad (136)$$

The forward probability distribution is then given by

$$p(n^B, m^A) = p_{\text{ini}}(m^A) p(n^B|m^A). \quad (137)$$

Next, let us define the reference probability distribution of the backward process. To derive Jarzynski equality, we take the initial state of the backward process as the canonical distribution

$$p_r(n^B) = \frac{1}{Z_B} e^{-\beta E_n^B}, \quad (138)$$

and the reference probability distribution is given by

$$p_r(m^A, n^B) = p_r(n^B) \tilde{p}(m^A|n^B), \quad (139)$$

where

$$\tilde{p}(m^A|n^B) = \left| \langle m^A | U^\dagger | n^B \rangle \right|^2 = p(n^B|m^A). \quad (140)$$

We derive the quantum Jarzynski equality by using the fact that the reference probability distribution is normalized:

$$\begin{aligned} 1 &= \sum_{m,n} p_r(m^A, n^B) \\ &= \sum_{m \notin X, n} p_r(m^A, n^B) + \sum_{m \in X, n} p_r(m^A, n^B), \end{aligned} \quad (141)$$

where X is a set of m^A 's satisfying $p_r(m^A, n^B) = 0$ if $p(m^A, n^B) = 0$, specifying ordinary irreversible processes. We note that the term with $n^B = \infty$ does not contribute to the sum in equation (141) because $p_r(m^A, n^B = \infty) = 0$. By taking the ratio between the forward probability and the backward probability, the

second term on the right-hand side of equation (141) is related to the ensemble average of the exponentiated work:

$$\begin{aligned}\sum_{m \in X, n} p_r(m^A, n^B) &= \sum_{m \in X, n} \frac{p_r(m^A, n^B)}{p(m^A, n^B)} p(m^A, n^B) \\ &= \sum_{m \in X, n} e^{\beta(E_m^A - E_n^B + F^B - F^A)} p(m^A, n^B) \\ &= \langle e^{\beta(W + \Delta F)} \rangle,\end{aligned}\quad (142)$$

where $W = E_m^A - E_n^B$ and $\Delta F = F^B - F^A$. By denoting the total probability of the absolutely irreversible process as $\lambda = \sum_{m \notin X, n} p_r(m^A, n^B)$, we can derive the quantum Jarzynski equality for a quantum piston:

$$\langle e^{\beta(W + \Delta F)} \rangle = 1 - \lambda. \quad (143)$$

Next, we will show that for finite ν , $\lambda = 0$ as discussed in [43]. However, if we take the limit of $\nu = \infty$, λ takes a nonzero value and the process is absolutely irreversible. The difference between the result given in [43] and ours is discussed in the appendix.

Note that the forward probability vanishes only when the initial probability distribution $p_{\text{ini}}(m^A)$ vanishes. This occurs only for $m^A = \infty$, meaning that X is given by a set of labels except $m^A = \infty$. (The contribution from $m^A = \infty$ is essential when we consider the case of $\nu = \infty$.) Then, λ is given by

$$\begin{aligned}\lambda &= \sum_n p_r(m^A = \infty, n^B) \\ &= \sum_n p_r(n^B) \tilde{p}(m^A = \infty | n^B).\end{aligned}\quad (144)$$

Let us first consider the case for finite ν , and show that there is no absolute irreversibility in this case. We see from equation (136) that the integral over x gives a nonzero value only when l is of the order of m ; otherwise the integrand oscillates rapidly due to the very large value of m^A . However, even if l is large, the integral over y vanishes due to the rapidly oscillating term, since n^B takes a finite value. We thus conclude $\tilde{p}(m^A = \infty | n^B) = 0$ and

$$\lambda = 0 \quad \text{for finite } \nu. \quad (145)$$

Now we move on to the interesting case of $\nu = \infty$ (see figure 3). We first consider the case in which m^A is small compared with ν (meaning that in the limit $\nu \rightarrow \infty$, $m^A \in X$). By using the stationary phase approximation, $p(n^B | m^A)$ is given by (see appendix for the detailed calculation)

$$p(n^B | m^A) = \left| \int_0^A \frac{2}{\sqrt{AB}} \sin\left(\frac{m\pi z}{A}\right) \sin\left(\frac{n\pi z}{B}\right) dz \right|^2. \quad (146)$$

For $m^A = \infty$ (or $m^A \notin X$), the conditional probability does not vanish since the oscillatory term $\sin(m\pi x/A)$ cancels with the term $\exp(-iM\nu x^2/(2A\hbar))$, and the conditional probability has a nonzero value:

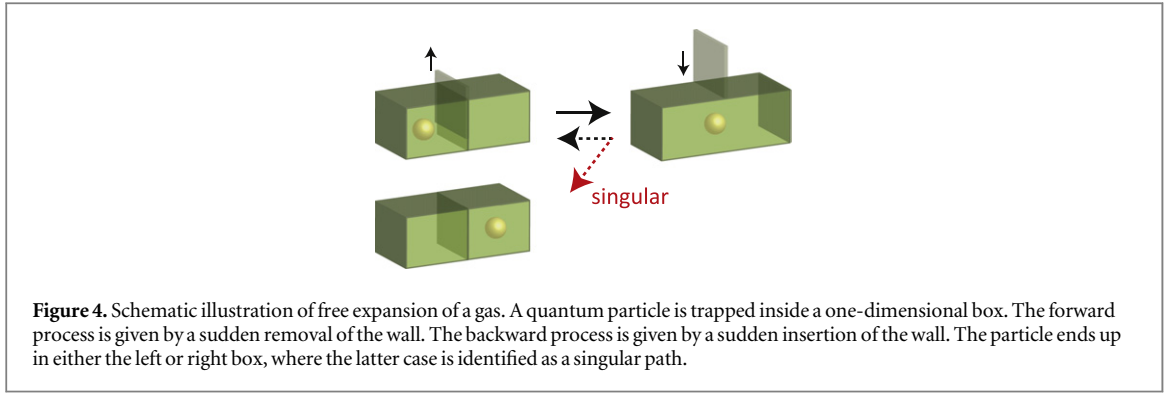
$$\lim_{\nu \rightarrow \infty} p(n^B | m^A = \infty) \neq 0. \quad (147)$$

We refer to [44] for the direct calculation of $p(n^B | m^A = \infty)$. Instead, we use the following relation to calculate λ :

$$\sum_m \tilde{p}(m^A | n^B) = 1, \quad (148)$$

which results from the unitary dynamics of the backward process. By using equation (146), we obtain

$$\begin{aligned}1 - \tilde{p}(m^A = \infty | n^B) &= \sum_{m \in X} p(m^A | n^B) \\ &= \frac{2}{B} \int_0^A dx \sin^2\left(\frac{n\pi x}{B}\right) \\ &= \frac{A}{B} - \frac{\sin\left(\frac{2\pi n A}{B}\right)}{2\pi n},\end{aligned}\quad (149)$$



and λ is given by

$$\begin{aligned}
 \lambda &= \sum_n p_r(n^B) \tilde{p}(m^A = \infty | n^B) \\
 &= \sum_n p_r(n^B) \left(1 - \frac{A}{B} - \frac{\sin\left(\frac{2\pi n A}{B}\right)}{2\pi n} \right) \\
 &= 1 - \frac{A}{B} - \sum_n \frac{\exp\left(-\frac{\beta n^2 \pi^2 \hbar^2}{2MB^2}\right) \sin\left(\frac{2\pi n A}{B}\right)}{Z^B 2\pi n}.
 \end{aligned} \tag{150}$$

We note that when we consider a sudden removal of the wall at $t = 0$ instead of pulling the wall (as illustrated in figure 4), the wave function of the initial state does not change. In this case, the exact form of the transition probability can be calculated with the sudden approximation:

$$\lim_{v \rightarrow \infty} U |m^A\rangle = |m^A\rangle, \tag{151}$$

$$\begin{aligned}
 p_{\text{sudden}}(n^B | m^A) &= \left| \langle n^B | m^A \rangle \right|^2 \\
 &= \left| \int_0^A \frac{2}{\sqrt{AB}} \sin\left(\frac{m\pi x}{A}\right) \sin\left(\frac{n\pi x}{B}\right) dx \right|^2,
 \end{aligned} \tag{152}$$

which is equal to the transition probability for $m \in X$ in equation (146). Therefore, the quantum Jarzynski equality for this process takes the same form as in equation (143) with λ given by equation (150).

Let us discuss the physical interpretation of the results. For $v \rightarrow \infty$ limit, the gas particle does not bounce at the wall in the forward process because the initial probability of having infinite energy ($m^A = \infty$) is zero. On the other hand, if we consider a reverse process describing an infinitely fast compression of a box, the particle located in the region $A < x < B$ will be pushed by the wall and the final energy of the particle is given by $E_{m=\infty}^A$. Since the state $|m^A = \infty\rangle$ is not occupied in the initial state, such paths are absolutely irreversible and the total probability of those backward paths is given by λ . If we model free expansion by a sudden removal of the wall, the forward process is exactly the same as the case of pulling the piston at an infinite speed. However, the backward process is described by the sudden insertion of the wall, which is different from the sudden compression of the piston. In this case, the absolutely irreversible process is given by the paths where the particle ends up in the region $A < x < B$, which was not occupied in the initial state. In these two cases, λ takes on the same value because they give the total probability of the backward process in which the particle is initially located in the region $A < x < B$.

4.2. Feedback control of qubit systems

In this subsection, we apply the quantum Jarzynski equality (112) to qubit systems. Let us prepare an initial state given by

$$\rho_{\text{ini}}^S = \rho_{\text{can}}^S = \frac{1}{1 + e^{-\beta V}} |0\rangle\langle 0|_S + \frac{e^{-\beta V}}{1 + e^{-\beta V}} |1\rangle\langle 1|_S, \tag{153}$$

where V is the energy difference between the two states $|0\rangle_S$ and $|1\rangle_S$.

Let us perform a projective measurement with respect to the basis set $\{|0\rangle_S, |1\rangle_S\}$. The probability p_k of obtaining the measurement outcome k is given by

$$p_0 = \frac{1}{1 + e^{-\beta V}}, \quad (154)$$

$$p_1 = \frac{e^{-\beta V}}{1 + e^{-\beta V}}, \quad (155)$$

and the postmeasurement state conditioned on the measurement outcome is given by a pure state

$$\rho^S(k) = |k\rangle\langle k|_S. \quad (156)$$

We can confirm that the acquired knowledge of the system $\langle I \rangle$ is equal to the Shannon entropy $H(p_k) = -\sum_k p_k \ln p_k$ calculated from the probability distribution of the measurement outcome:

$$\begin{aligned} \langle I \rangle &= S(\rho_{\text{ini}}^S) - \sum_k p_k S(\rho^S(k)) \\ &= H(p_k) = \beta V \cdot \frac{e^{-\beta V}}{1 + e^{-\beta V}} + \ln(1 + e^{-\beta V}). \end{aligned} \quad (157)$$

Note that the state of the system conditioned on the measurement outcome is less mixed for a greater value of $\langle I \rangle$. In this case, the postmeasurement state is given by a pure state and our knowledge of the state has increased by the maximum amount $\langle I \rangle = H(p_k)$ due to the measurement.

Depending on the measurement outcome, we perform the following feedback control which has the effect of flipping the state if the post-measurement state is $|1\rangle_S$:

$$U_0 = 1, \quad U_1 = |0\rangle\langle 1|_S + |1\rangle\langle 0|_S. \quad (158)$$

After the feedback control, we obtain

$$\rho_{\text{fb}}^S(k) = U_k |k\rangle\langle k| U_k^\dagger = |0\rangle\langle 0|_S, \quad (159)$$

which is independent of the measurement outcome k . The averaged density matrix is given by $\rho_{\text{fb}}^S = \sum_k p_k \rho_{\text{fb}}^S(k) = |0\rangle\langle 0|_S$. The energy change of the system during the feedback is given by

$$\Delta E_{\text{fb}} = -V \frac{e^{-\beta V}}{1 + e^{-\beta V}}. \quad (160)$$

We model the isothermal expansion process by introducing $N + 1$ different heat baths, each of which is composed of a qubit, as schematically illustrated in figure 5. A similar model is discussed in [56]. The Hamiltonian of each heat bath is given by

$$H^{B_n} = (E_0 - n\Delta V) |1\rangle\langle 1|_{B_n}, \quad n = 0, \dots, N, \quad (161)$$

where $E_0 = N\Delta V + V$ is the energy difference between two states of the zeroth heat bath. The initial state of the entire heat bath is given by the tensor product of the canonical distributions:

$$\rho_{\text{can}}^B = \bigotimes_n \left(\frac{1}{1 + e^{-\beta(E_0 - n\Delta V)}} |0\rangle\langle 0|_{B_n} + \frac{e^{-\beta(E_0 - n\Delta V)}}{1 + e^{-\beta(E_0 - n\Delta V)}} |1\rangle\langle 1|_{B_n} \right). \quad (162)$$

We consider the following $N + 1$ steps of the protocol to thermalize the system.

(a) We quench the Hamiltonian of the system so that the energy difference of the system is changed from E to E_0 . Note that this process preserves the energy of the system since the excited state $|1\rangle_S$ is not populated during this process. Next, we perform the following unitary transformation between the system and B_0 :

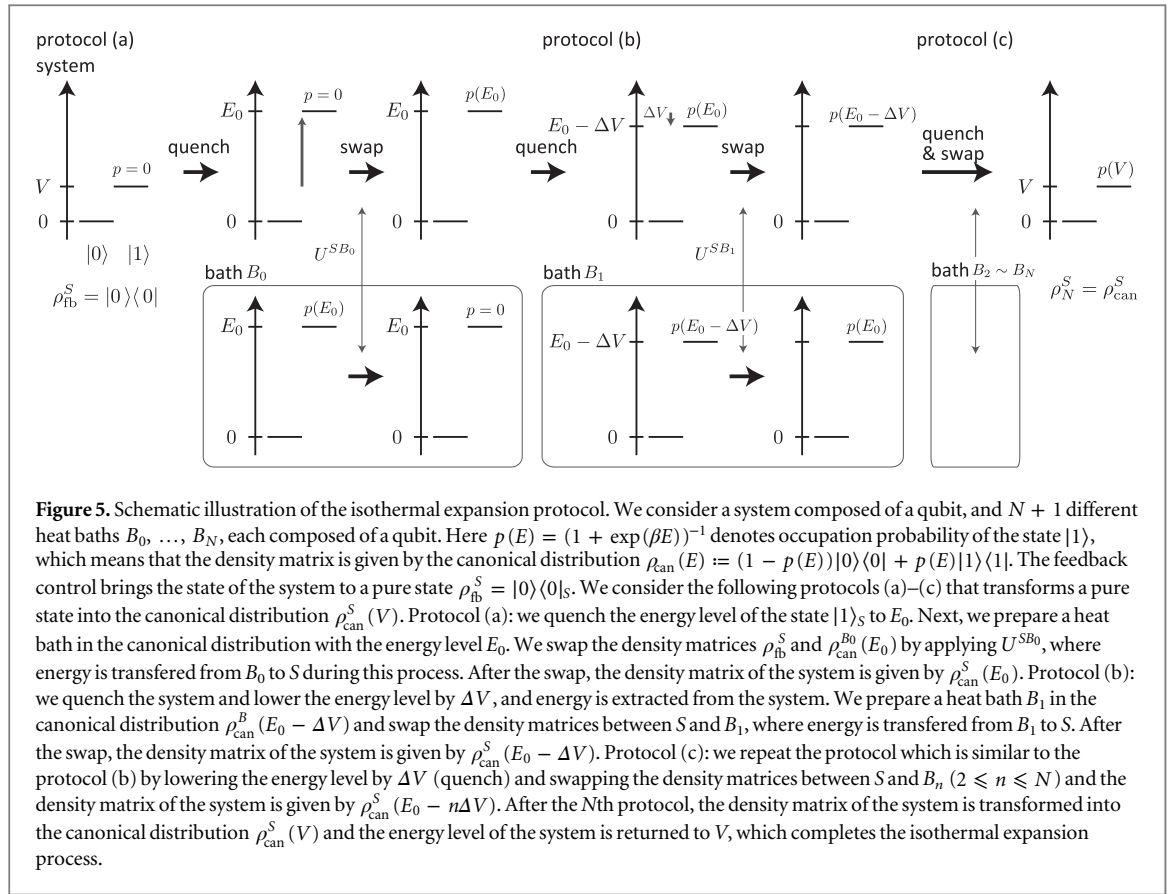
$$U^{SB_0} = |0\rangle\langle 0|_S \otimes |0\rangle\langle 0|_{B_0} + |1\rangle\langle 0|_S \otimes |0\rangle\langle 1|_{B_0} + |0\rangle\langle 1|_S \otimes |1\rangle\langle 0|_{B_0} + |1\rangle\langle 1|_S \otimes |1\rangle\langle 1|_{B_0}. \quad (163)$$

This transformation swaps the populations between S and B_0 :

$$\begin{aligned} \rho_0^{SB_0} &= U^{SB_0} \left(|0\rangle\langle 0|_S \otimes \left(\frac{1}{1 + e^{-\beta E_0}} |0\rangle\langle 0|_{B_0} + \frac{e^{-\beta E_0}}{1 + e^{-\beta E_0}} |1\rangle\langle 1|_{B_0} \right) \right) U^{\dagger SB_0} \\ &= \left(\frac{1}{1 + e^{-\beta E_0}} |0\rangle\langle 0|_S + \frac{e^{-\beta E_0}}{1 + e^{-\beta E_0}} |1\rangle\langle 1|_S \right) \otimes |0\rangle\langle 0|_{B_0}. \end{aligned} \quad (164)$$

During this process, the energy flows from B_0 to S . The energy change $\Delta E = E_{\text{fin}} - E_{\text{ini}}$ of B_0 and that of S can be explicitly calculated as

$$\Delta E^{B_0} = -E_0 \cdot \frac{e^{-\beta E_0}}{1 + e^{-\beta E_0}}, \quad (165)$$



$$\Delta E_0^S = E_0 \cdot \frac{e^{-\beta E_0}}{1 + e^{-\beta E_0}}. \quad (166)$$

We note that the total energy change is zero: $\Delta E^{B_0} + \Delta E_0^S = 0$.

(b) We quench the Hamiltonian of the system so that the energy difference is changed from E_0 to $E_0 - \Delta V$. During this process, the energy change of the system is given by

$$\Delta E_{q,1}^S = -\Delta V \cdot \frac{e^{-\beta E_0}}{1 + e^{-\beta E_0}}. \quad (167)$$

Next, we let S interact with B_1 via the unitary transformation which has the same form as in equation (163). After the swap, the density matrix is given by

$$\begin{aligned} \rho_1^{SB_1} &= U^{SB_1} \left(\rho_0^S \otimes \rho_{\text{can}}^{B_1} \right) U^{\dagger SB_1} \\ &= \left(\frac{1}{1 + e^{-\beta(E_0 - \Delta V)}} |0\rangle\langle 0|_S + \frac{e^{-\beta(E_0 - \Delta V)}}{1 + e^{-\beta(E_0 - \Delta V)}} |1\rangle\langle 1|_S \right) \\ &\quad \otimes \left(\frac{1}{1 + e^{-\beta E_0}} |0\rangle\langle 0|_{B_1} + \frac{e^{-\beta E_0}}{1 + e^{-\beta E_0}} |1\rangle\langle 1|_{B_1} \right), \end{aligned} \quad (168)$$

and the energy changes of B_1 and S are given by

$$\Delta E^{B_1} = -(E_0 - \Delta V) \cdot \left(\frac{e^{-\beta(E_0 - \Delta V)}}{1 + e^{-\beta(E_0 - \Delta V)}} - \frac{e^{-\beta E_0}}{1 + e^{-\beta E_0}} \right) = -\Delta E_1^S. \quad (169)$$

(c) For the n th step ($2 \leq n \leq N$), we quench the Hamiltonian of the system so that the energy difference is changed from $E_0 - (n-1)\Delta V$ to $E_0 - n\Delta V$. During this process, the energy change of the system is given by

$$\Delta E_{q,n}^S = -\Delta V \frac{e^{-\beta(E_0 - (n-1)\Delta V)}}{1 + e^{-\beta(E_0 - (n-1)\Delta V)}}. \quad (170)$$

Next, we interact S and B_n using the same unitary transformation as in equation (163). After the swap, the density matrix is given by

$$\begin{aligned}
\rho_n^{SB_n} &= U^{SB_n} \left(\rho_{n-1}^S \otimes \rho_{\text{can}}^{B_n} \right) U^{\dagger SB_n} \\
&= \left(\frac{1}{1 + e^{-\beta(E_0 - n\Delta V)}} |0\rangle\langle 0|_S + \frac{e^{-\beta(E_0 - n\Delta V)}}{1 + e^{-\beta(E_0 - n\Delta V)}} |1\rangle\langle 1|_S \right) \\
&\quad \otimes \left(\frac{1}{1 + e^{-\beta(E_0 - (n-1)\Delta V)}} |0\rangle\langle 0|_{B_n} + \frac{e^{-\beta(E_0 - (n-1)\Delta V)}}{1 + e^{-\beta(E_0 - (n-1)\Delta V)}} |1\rangle\langle 1|_{B_n} \right),
\end{aligned} \tag{171}$$

and the energy change is given by

$$\Delta E^{B_n} = -(E_0 - n\Delta V) \cdot \left(\frac{e^{-\beta(E_0 - n\Delta V)}}{1 + e^{-\beta(E_0 - n\Delta V)}} - \frac{e^{-\beta(E_0 - (n-1)\Delta V)}}{1 + e^{-\beta(E_0 - (n-1)\Delta V)}} \right) = -\Delta E_n^S. \tag{172}$$

After the N th step, the system returns to the canonical distribution, which is the final state of this protocol:

$$\rho_N^S = \rho_{\text{can}}^S. \tag{173}$$

We use the short-hand notation

$$|h\rangle_B = |h_0\rangle_{B_0} \otimes |h_1\rangle_{B_1} \otimes \cdots \otimes |h_N\rangle_{B_N}, \tag{174}$$

where h_n takes the value 0 or 1, and $|h_n\rangle_{B_n}$ describes the energy eigenstate of the n th heat bath. We also use the notation $U^{SB} = U^{SB_N} U^{SB_{N-1}} \cdots U^{SB_0}$, which is the total unitary operation performed on the total system during the isothermal expansion process. Now we explicitly calculate the left-hand side of equation (112) as follows:⁵

$$\langle e^{\beta W - I} \rangle = \sum_{h,j,k,z} \left| \langle z|_S \otimes \langle j|_B U^{SB} U_k^S |k\rangle_S \otimes |h\rangle_B \right|^2 \frac{e^{-\beta E^B(h)}}{Z_B} \frac{e^{-\beta E^S(k)}}{Z_S} e^{\beta W(k,z,h,j) - I(k)}, \tag{175}$$

where

$$\frac{e^{-\beta E^B(h)}}{Z_B} = \frac{e^{-\beta E^{B_0}(h_0)}}{Z_{B_0}} \cdot \frac{e^{-\beta E^{B_1}(h_1)}}{Z_{B_1}} \cdots \frac{e^{-\beta E^{B_N}(h_N)}}{Z_{B_N}}, \tag{176}$$

and

$$E^{B_n}(h_n) = \begin{cases} 0 & h_n = 0; \\ E_0 - n\Delta V & h_n = 1, \end{cases} \tag{177}$$

$$E^S(k) = \begin{cases} 0 & k = 0; \\ V & k = 1. \end{cases} \tag{178}$$

Noting that the work in equation (175) is given by

$$W(h, j, k, z) = E^S(k) - E^S(z) + E^B(h) - E^B(j), \tag{179}$$

we can further calculate equation (175) and obtain

$$\begin{aligned}
\langle e^{\beta W - I} \rangle &= \sum_{h,j,k,z} p_k \left| \langle z|_S \otimes \langle j|_B U^{SB} U_k^S |k\rangle_S \otimes |h\rangle_B \right|^2 \frac{e^{-\beta E^B(j)}}{Z_B} \frac{e^{-\beta E^S(z)}}{Z_S} \\
&= \sum_k p_k \langle k|_S U_k^{\dagger S} \text{Tr}_B \left[U^{\dagger SB} \left(\rho_{\text{can}}^S \otimes \rho_{\text{can}}^B \right) U^{SB} \right] U_k^S |k\rangle_S.
\end{aligned} \tag{180}$$

Due to the reverse protocol given above, the density matrix of the system returns to the state

$$\text{Tr}_B \left[U^{\dagger SB} \left(\rho_{\text{can}}^S \otimes \rho_{\text{can}}^B \right) U^{SB} \right] = \frac{1}{1 + e^{-\beta E_0}} |0\rangle\langle 0|_S + \frac{e^{-\beta E_0}}{1 + e^{-\beta E_0}} |1\rangle\langle 1|_S. \tag{181}$$

Using equation (181), we have an explicit form of equation (175):

$$\langle e^{\beta W - I} \rangle = \frac{1}{1 + e^{-\beta E_0}} = 1 - \lambda^{\text{fb}}, \tag{182}$$

where λ^{fb} is the total probability of the backward process not returning to the postmeasurement state $|k\rangle_S$ as given in equation (104):

⁵ While we derive equation (112) for a single heat bath, a generalization to multiple heat baths is straightforward.

$$\begin{aligned}\lambda^{\text{fb}} &= \sum_{k \neq y} p_k \langle y | s \rangle U_k^{\dagger S} \text{Tr}_B \left[U^{\dagger SB} \left(\rho_{\text{can}}^S \otimes \rho_{\text{can}}^B \right) U^{SB} \right] U_k^S |y\rangle_s \\ &= \frac{e^{-\beta E_0}}{1 + e^{-\beta E_0}}.\end{aligned}\quad (183)$$

Using Jensen's inequality and equation (182), we can derive the upper bound on the extractable work from the system via feedback control, that is

$$\langle W \rangle \leq k_B T \langle I \rangle + k_B T \ln(1 - \lambda^{\text{fb}}). \quad (184)$$

The right-hand side of (184) can be explicitly calculated as

$$k_B T \left(\langle I \rangle + \ln(1 - \lambda^{\text{fb}}) \right) = V \frac{e^{-\beta V}}{1 + e^{-\beta V}} + k_B T \ln \frac{1 + e^{-\beta V}}{1 + e^{-\beta E_0}}. \quad (185)$$

We can also calculate the work defined in equation (114):

$$\langle W \rangle = -\Delta E^S - \Delta E^B, \quad (186)$$

where ΔE^S and ΔE^B are the total energy change of S and B , respectively. As the system returns to the initial state at the end of the protocol, $\Delta E^S = 0$. The total energy change of the heat bath is given by

$$-\Delta E^B = -\sum_n \Delta E^{B_n} = \sum_{n=0}^{N-1} \Delta V \frac{e^{-\beta(E_0 - n\Delta V)}}{1 + e^{-\beta(E_0 - n\Delta V)}} + V \frac{e^{-\beta V}}{1 + e^{-\beta V}}. \quad (187)$$

We can also interpret work as the energy that is extracted in the course of the quench process during (a)–(c) plus the extracted energy during the flipping process of the feedback control, that is

$$\begin{aligned}\langle W \rangle &= -\sum_n \Delta E_{q,n}^S - \Delta E_{\text{fb}}^S \\ &= \sum_{n=0}^{N-1} \Delta V \frac{e^{-\beta(E_0 - n\Delta V)}}{1 + e^{-\beta(E_0 - n\Delta V)}} + V \frac{e^{-\beta V}}{1 + e^{-\beta V}},\end{aligned}\quad (188)$$

which gives the same amount of work compared with the extracted work defined in equation (186). As we fix $E_0 = N\Delta V + V$ and take the limit $\Delta V \rightarrow 0$ (and $N \rightarrow \infty$), the right-hand side of equation (187) reaches

$$-\Delta E^B|_{\Delta V \rightarrow 0} = V \frac{e^{-\beta V}}{1 + e^{-\beta V}} + k_B T \ln \frac{1 + e^{-\beta V}}{1 + e^{-\beta E_0}}. \quad (189)$$

Since $\Delta E^B \leq \Delta E^B|_{\Delta V \rightarrow 0}$, inequality (184) is valid and the equality condition is achieved in the limit of $\Delta V \rightarrow 0$ and $N \rightarrow \infty$:

$$\langle W \rangle = k_B T \langle I \rangle + k_B T \ln(1 - \lambda^{\text{fb}}), \quad (\Delta V \rightarrow 0). \quad (190)$$

If we consider a finite ΔV , the density matrix of S jumps from ρ_n^S to ρ_{n+1}^S during the process, causing dissipation. This dissipation is due to the ordinary irreversibility of the process and not due to absolute irreversibility, since only the relative weights of two states $|0\rangle_S$ and $|1\rangle_S$ are changed during the protocols between (a) and (c).

The effect of absolute irreversibility depends on the parameter E_0 for this model, since the protocol (a) brings the pure state $|0\rangle_S$ of the postmeasurement state into a thermal state. If we take the limit $E_0 \rightarrow \infty$, we have no absolute irreversibility ($\lambda^{\text{fb}} = 0$) and the backward process corresponding to the forward protocol (a) makes the density matrix return to $|0\rangle_S$. In this limit, one can extract work up to the amount commensurate with the amount of information $\langle I \rangle$ obtained via measurement:

$$\langle W \rangle \leq k_B T \langle I \rangle \quad (E_0 \rightarrow \infty), \quad (191)$$

and the equality is again in the $\Delta V \rightarrow 0$ limit:

$$\langle W \rangle = k_B T \langle I \rangle \quad (\Delta V \rightarrow 0, E_0 \rightarrow \infty), \quad (192)$$

where the acquired information is fully utilized to extract work. The protocol we consider (in the limit of $\Delta V \rightarrow 0$ and $E_0 \rightarrow \infty$, so that the system interacts with infinitely many heat baths) gives a quasi-static process of isothermal expansion of system in the sense that

$$\Delta S^S = \beta Q \quad (193)$$

is achieved, where $\Delta S^S = S(\rho_{\text{can}}^S) - S(|0\rangle\langle 0|_S)$ gives a change in the von Neumann entropy of the system during the isothermal expansion process (a)–(c) and

$$Q = \Delta E^B \quad (194)$$

is the heat taken from the heat baths. We can relate the energy change of the heat baths to heat because the total change in the von Neumann entropy of the heat bath satisfies the thermodynamic relation

$$\Delta S^B = \sum_n S(\rho_n^{B_n}) - S(\rho_{\text{can}}^B) = -\beta \Delta E^B. \quad (195)$$

As a result, equation (192) is satisfied and the reduced entropy of the system via feedback is fully converted into work by this quasi-static process. This result is to be compared with the classical single-particle Szilard engine that achieves equation (192) via quasistatic isothermal expansion of the box [58].

Next, let us consider the opposite limit of $E_0 = V$. In this case, we do not quench the energy level of the system. We only attach a single heat bath, letting the postmeasurement state of the system transform into a thermal state by a single jump (the protocol (a)). The effect of absolute irreversibility is maximal in this limit:

$$\lambda^{\text{fb}}|_{E_0=V} = \frac{e^{-\beta V}}{1 + e^{-\beta V}} \geq \frac{e^{-\beta E_0}}{1 + e^{-\beta E_0}} = \lambda^{\text{fb}}, \quad (196)$$

and the work gain takes the smallest value

$$\langle W \rangle_{E_0=V} = V \frac{e^{-\beta V}}{1 + e^{-\beta V}}. \quad (197)$$

5. Conclusion

We have derived the quantum fluctuation theorem (38) and Jarzynski equality (43) in the presence of absolutely irreversible processes, where the density matrix of the backward process does not return to the subspace spanned by the eigenvectors that have nonzero overlap with the initial density matrix. We have also derived equalities for feedback and measurement processes (100) and (101). The absolute irreversibility limits the work gain via (inefficient) feedback control and also gives additional entropy production due to the projection on the memory (102) and (103). The latter fact means that the dissipation by quantum decoherence can be qualitatively characterized by absolute irreversibility. We have also discussed a model of the quantum piston and the feedback control on a qubit system to illustrate the obtained nonequilibrium equalities.

Acknowledgments

This work was supported by KAKENHI Grant No. 26287088 from the Japan Society for the Promotion of Science, a Grant-in-Aid for Scientific Research on Innovation Areas ‘Topological Quantum Phenomena’ (KAKENHI Grant No. 22103005), the Photon Frontier Network Program from MEXT of Japan, and the Mitsubishi Foundation. KF acknowledges support from JSPS (Grant No. 254105) and through Advanced Leading Graduate Course for Photon Science (ALPS). YM was supported by the Japan Society for the Promotion of Science through Program for Leading Graduate Schools (MERIT). KF thanks Adolfo del Campo, Jordan M Horowitz and Yuto Ashida for fruitful discussions.

Appendix. Quantum pistons

1. Derivation of equation (146)

We derive equation (146) for large ν and small m . Note that the integral over x in equation (136) vanishes for small l due to the rapidly oscillating terms. If l is the order of ν , the integral can be approximated by using the stationary phase approximation:

$$\frac{2}{A} \int_0^A e^{-\frac{iM\nu x^2}{2A\hbar}} \sin\left(\frac{l\pi x}{A}\right) \sin\left(\frac{m\pi x}{A}\right) dx \simeq -\frac{i}{A} e^{\frac{i\pi^2 l^2 \hbar}{2AM\nu}} \sin\left(\frac{lm\pi^2 \hbar}{AM\nu}\right) \sqrt{\frac{2\pi A \hbar}{M\nu}} e^{-\frac{i\pi}{4}}, \quad (198)$$

where $(l\pi\hbar)/M\nu \leq A$. We can also approximate the integral for y :

$$\frac{2}{B} \int_0^B e^{\frac{iM\nu y^2}{2B\hbar}} \sin\left(\frac{l\pi y}{A}\right) \sin\left(\frac{n\pi y}{B}\right) dy \simeq \frac{i}{B} e^{-\frac{i\pi^2 l^2 \hbar}{2BM\nu}} \sin\left(\frac{ln\pi^2 \hbar}{BM\nu}\right) \sqrt{\frac{2\pi B \hbar}{M\nu}} e^{\frac{i\pi}{4}}, \quad (199)$$

where $(l\pi\hbar)/M\nu \leq B$. Now $P(n|m)$ takes the form

$$p(n^B|m^A) = \left| \sum_l \frac{\hbar}{M\nu} \frac{2\pi}{\sqrt{AB}} \sin\left(\frac{lm\pi^2 \hbar}{AM\nu}\right) \sin\left(\frac{ln\pi^2 \hbar}{BM\nu}\right) \right|^2, \quad (200)$$

where $(l\pi\hbar)/Mv \leq A \leq B$. Note that we have extended the summation to small l , since they give a negligible contribution when we take the limit of $v \rightarrow \infty$. By taking the $v \rightarrow \infty$ limit, we replace the sum with the integral, i.e.

$$z = \frac{l\pi\hbar}{Mv}, \quad dz = \frac{\pi\hbar}{Mv}. \quad (201)$$

Then, combining equations (200) and (201), we obtain the desired result (146).

2. Comparison with previous work

In this subsection, we compare our results with the results derived in [43]. In [43], the authors derived the quantum Jarzynski equality for some finite v and obtained equation (143) with $\lambda = 0$. They also discussed the different order of taking the limits for the velocity v and the number N of samplings of the ensemble average. To be specific, let us write the left-hand side of the Jarzynski equality with a fixed velocity v and take the ensemble average for some finite number N :

$$f_{N,v} := \langle \exp(\beta(W + \Delta F)) \rangle_{N,v}. \quad (202)$$

The authors of [43] discussed the validity of the Jarzynski equality for free expansion by taking the infinite limit of v after taking N to infinity: $\lim_{v \rightarrow \infty} \lim_{N \rightarrow \infty} f_{N,v} = 1$. However, this operation implicitly assumes that a particle has to bounce at the infinitely fast moving wall. Therefore, it is questionable to interpret this process as a free expansion. In a mathematical sense, this limit corresponds to taking the following limits:

$$\begin{aligned} \lim_{v \rightarrow \infty} \lim_{N \rightarrow \infty} f_{N,v} &= \lim_{v \rightarrow \infty} \sum_{m \in X,n} p_r(m^A, n^B) + \lim_{v \rightarrow \infty} \sum_{m \notin X,n} p_{\text{ini}}(m^A) e^{\beta(E_m^A - \Delta F^A)} p_r(m^A, n^B) \\ &= \lim_{v \rightarrow \infty} \sum_{m \in X,n} p_r(m^A, n^B) + \lim_{v \rightarrow \infty} \sum_{m \notin M,n} \frac{p_{\text{ini}}(m^A)}{p_{\text{ini}}(m^A)} p_r(m^A, n^B). \end{aligned} \quad (203)$$

Here special care should be taken to the second term on the right-hand side of equation (203) because the $v \rightarrow \infty$ limit means

$$\lim_{v \rightarrow \infty} \sum_{m \notin X,n} \frac{p_{\text{ini}}(m^A)}{p_{\text{ini}}(m^A)} p_r(m^A, n^B) = \sum_n \frac{0}{0} p_r(m^A = \infty, n^B). \quad (204)$$

Only if we assume $0/0 = 1$, the right-hand side of equation (203) is equal to unity and the quantum Jarzynski equality is obtained:

$$\lim_{v \rightarrow \infty} \lim_{N \rightarrow \infty} f_{N,v} = 1. \quad (205)$$

On the other hand, our result in equation (143) is obtained by taking the other ordering of the limits:

$$\lim_{N \rightarrow \infty} \lim_{v \rightarrow \infty} f_{N,v} = 1 - \lambda, \quad (206)$$

which describes a more physical situation of free expansion compared with that in equation (205).

References

- [1] Evans D J, Cohen E G D and Morriss G P 1993 *Phys. Rev. Lett.* **71** 2401–4
- [2] Gallavotti G and Cohen E G D 1995 *Phys. Rev. Lett.* **74** 2694–7
- [3] Crooks G E 1999 *Phys. Rev. E* **60** 2721–6
- [4] Jarzynski C 1997 *Phys. Rev. Lett.* **78** 2690
- [5] Jarzynski C 1997 *Phys. Rev. E* **56** 5018
- [6] Seifert U 2005 *Phys. Rev. Lett.* **95** 040602
- [7] Evans D J and Searles D J 2002 *Adv. Phys.* **51** 1529
- [8] Seifert U 2012 *Rep. Prog. Phys.* **75** 126001
- [9] Schuler S, Speck T, Tietz C, Wrachtrup J and Seifert U 2012 *Phys. Rev. Lett.* **94** 180602
- [10] Tietz C, Schuler S, Speck T, Seifert U and Wrachtrup J 2006 *Phys. Rev. Lett.* **97** 050602
- [11] Hatano T and Sasa S 2001 *Phys. Rev. Lett.* **86** 3463
- [12] Trepagnier E H, Jarzynski C, Ritort F, Crooks G E, Bustamante C J and Liphardt J 2004 *Proc. Natl Acad. Sci. USA* **101** 15038
- [13] Tasaki H 1999 arXiv: [cond-mat/0009244](https://arxiv.org/abs/cond-mat/0009244)
- [14] Kurchan J 2000 arXiv: [cond-mat/0007360](https://arxiv.org/abs/cond-mat/0007360)
- [15] Esposito M, Harbola U and Mukamel S 2009 *Rev. Mod. Phys.* **81** 1665
- [16] Monnai T 2005 *Phys. Rev. E* **72** 027102
- [17] Talkner P, Lutz E and Hänggi P 2007 *Phys. Rev. E* **75** 050102
- [18] Talkner P and Hänggi P 2007 *J. Phys. A: Math. Theor.* **40** F569
- [19] Campisi M, Talkner P and Hänggi P 2009 *Phys. Rev. Lett.* **102** 210401
- [20] Campisi M, Talkner P and Hänggi P 2010 *Phys. Rev. Lett.* **105** 140601
- [21] Campisi M, Hänggi P and Talkner P 2011 *Rev. Mod. Phys.* **83** 771

- [22] Horowitz J M 2012 *Phys. Rev. E* **85** 031110
- [23] Horowitz J M and Parrondo J M 2013 *New J. Phys.* **15** 085028
- [24] Chetrite R and Mallick K 2012 *J. Stat. Phys.* **148** 480–501
- [25] Subasi Y and Hu B L 2012 *Phys. Rev. E* **85** 011112
- [26] Talkner P, Campisi M and Hänggi P 2009 *J. Stat. Mech.* **2009** P02025
- [27] Crooks G E 2008 *J. Stat. Mech.* **2008** P10023
- [28] Sagawa T and Ueda M 2010 *Phys. Rev. Lett.* **104** 090602
- [29] Sagawa T and Ueda M 2012 *Phys. Rev. Lett.* **109** 180602
- [30] Sagawa T 2013 *Thermodynamics of Information Processing in Small Systems* (Berlin: Springer)
- [31] Morikuni M and Tasaki H 2011 *J. Stat. Phys.* **143** 1
- [32] Funo K, Watanabe Y and Ueda M 2013 *Phys. Rev. E* **88** 052121
- [33] Collin D, Ritort F, Jarzynski C, Smith S B, Tinoco I Jr and Bustamante C 2005 *Nature* **437** 231–4
- [34] Liphardt J, Dumont S, Smith S B, Tinoco I Jr and Bustamante C 2002 *Science* **296** 1832
- [35] An S, Zhang J, Um M, Lv D, Lu Y, Zhang J, Yin Z, Quan H T and Kim K 2015 *Nat. Phys.* **11** 193
- [36] Maxwell J C 1871 *Theory of Heat* (London: Appleton)
- [37] Toyabe S, Sagawa T, Ueda M, Muneyuki E and Sano M 2010 *Nat. Phys.* **6** 988
- [38] Sung J 2005 arXiv: [cond-mat/0506214](#)
- [39] Lua R C and Grosberg A Y 2005 *J. Phys. Chem. B* **109** 6805
- [40] Gross D H E 2005 arXiv: [cond-mat/0508721](#)
- [41] Jarzynski C 2005 arXiv: [cond-mat/0509344](#)
- [42] Gross D H E 2005 arXiv: [cond-mat/0509648](#)
- [43] Quan H T and Jarzynski C 2012 *Phys. Rev. E* **85** 031102
- [44] Dodonov V V, Klimov A B and Nikonov D E 1993 *J. Math. Phys.* **34** 3391
- [45] Ashida Y, Funo K, Murashita Y and Ueda M 2014 *Phys. Rev. E* **90** 052125
- [46] Murashita Y, Funo K and Ueda M 2014 *Phys. Rev. E* **90** 042110
- [47] Sagawa T and Ueda M 2008 *Phys. Rev. Lett.* **100** 080403
- [48] Sagawa T and Ueda M 2009 *Phys. Rev. Lett.* **102** 250602
- [49] Sagawa T 2012 *Lectures on Quantum Computing, Thermodynamics, and Statistical Physics* ed M Nakahara and S Tanaka (Singapore: World Scientific)
- [50] Santos M F and Carvalho A R R 2011 *Europhys. Lett.* **94** 64003
- [51] Groenewold H J 1971 *Int. J. Theor. Phys.* **4** 327
- [52] Ozawa M 1986 *J. Math. Phys.* **27** 759
- [53] Nielsen M A and Chuang I L 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
- [54] Halmos P R 1974 *Measure Theory* (Berlin: Springer) p 134, 182
- [55] Bartle R G 1995 *The Elements of Integration and Lebesgue Measure* (New York: Wiley) p 88
- [56] Skrzypczyk P, Short A J and Popescu S 2013 arXiv: [1302.2811](#)
- [57] Horowitz J M and Parrondo J M R 2011 *Europhys. Lett.* **95** 10005
- [58] Szilard L 1929 *Z. Phys.* **53** 840