npj | quantum information

Article

Published in partnership with The University of New South Wales

https://doi.org/10.1038/s41534-024-00922-w

Universal validity of the second law of information thermodynamics

Shintaro Minagawa [1] ✉, M. Hamed Mohammady [2,3] ✉, Kenta Sakai[1] ✉, Kohtaro Kato[1] ✉ & Francesco Buscemi[1] ✉

Adiabatic measurements, followed by feedback and erasure protocols, have often been considered as a model to embody Maxwell's Demon paradox and to study the interplay between thermodynamics and information processing. Such studies have led to the conclusion, now widely accepted in the community, that Maxwell's Demon and the second law of thermodynamics can peacefully coexist because any gain provided by the demon must be offset by the cost of performing the measurement and resetting the demon's memory to its initial state. Statements of this kind are collectively referred to as *second laws of information thermodynamics* and have recently been extended to include quantum theoretical scenarios. However, previous studies in this direction have made several assumptions, particularly about the feedback process and the demon's memory readout, and thus arrived at statements that are not universally applicable and whose range of validity is not clear. In this work, we fill this gap by precisely characterizing the full range of quantum feedback control and erasure protocols that are overall consistent with the second law of thermodynamics. This leads us to conclude that the second law of information thermodynamics is indeed *universal*: it must hold for any quantum feedback control and erasure protocol, regardless of the measurement process involved, as long as the protocol is overall compatible with thermodynamics. Our comprehensive analysis not only encompasses new scenarios but also retrieves previous ones, doing so with fewer assumptions. This simplification contributes to a clearer understanding of the theory.

The problem of consistency between the second law of thermodynamics and information processing has been at the center of one of the longest running debates in the history of modern physics, ever since Maxwell conjured up his famous demon[1]. A widely accepted solution to Maxwell's paradox is that consistency with the second law of thermodynamics is recovered by taking into account the work cost for measurement and erasure, i.e., the resetting of the demon's memory to its initial state[2–8]. These ideas, bridging thermodynamics with information theory, are nowadays collectively referred to as *information thermodynamics*[9,10].

In this context, and including a quantum theoretical scenario, Sagawa and Ueda, in a series of celebrated papers[11–13], derived an achievable upper bound for the work extracted by feedback control and showed that the conventional second law can, in general, be violated from the viewpoint of the system alone, but such a violation is exactly compensated by the cost of implementing the controlling measurement and resetting the memory. Such a tradeoff relation is what they call *the second law of information thermodynamics (ITh)*.

Unfortunately, despite their importance, the balance equations established in refs. 11–13 rely on several mutually inconsistent assumptions that lack a direct operational interpretation. Moreover, these works only discuss *sufficient* conditions for the validity of such balance equations. While some generalizations and refinements have been proposed[14–22], the demon's memory readout process is always limited to *ideal projective measurements*. Besides being unrealistic in practice, such an assumption is problematic *in principle*: since the demon's memory enters directly into the thermodynamic balance, the process acting on it must be treated *in full generality*, lest we obtain statements of limited scope. As a result, a comprehensive characterization of the validity range of the second law of ITh remains elusive, and it is unclear under what conditions the second law of ITh holds. In fact, at the time of writing, it is not even clear whether the second law of

[1]Graduate School of Informatics, Nagoya University, Furo-cho, Chikusa-Ku, Nagoya, 464-8601, Japan. [2]QuIC, École Polytechnique de Bruxelles, CP 165/59, Université Libre de Bruxelles, 1050 Brussels, Belgium. [3]RCQI, Institute of Physics, Slovak Academy of Sciences, Dúbravská cesta 9, Bratislava, 84511, Slovakia. ✉e-mail: minagawa.shintaro@nagoya-u.jp; m.hamed.mohammady@savba.sk; sakai.kenta_32@nagoya-u.jp; kokato@i.nagoya-u.ac.jp; buscemi@nagoya-u.jp

ITh should be considered a universal law or not, and what its logical status is with respect to the conventional second law of thermodynamics.

Our paper addresses this gap by adopting a top-down approach. Instead of attempting to *derive* the second law from assumptions with unclear logical necessity, we initiate from a purely information-theoretic framework and obtain balance equations that hold *for any measurement and isothermal feedback process*, in particular including any readout mechanism, and subsequently *impose* the second law of phenomenological thermodynamics as a constraint. This approach, which follows that used by von Neumann to derive his entropy's equation[23,24], allows us to determine exactly (in terms of sufficient *and necessary* conditions) how far feedback control and erasure protocols can be generalized while remaining overall consistent with the second law. We are then able to demonstrate the universal validity of the second law of ITh in general feedback control and erasure protocols: as long as such a protocol is compatible with the second law of phenomenological thermodynamics, it must also satisfy the second law of ITh, regardless of the measurement and feedback process involved.

A quantity that plays a crucial role in our analysis is the Groenewold–Ozawa information gain[25,26]: while previous works[14–16,19,27,28] have also provided it with a thermodynamic interpretation—even in situations when it takes negative values—our balance equations show that such interpretation holds in complete generality.

Results

The minimum scenario required to discuss Maxwell's paradox and feedback control protocols in full generality, but without oversimplifications, comprises five systems, as shown in Fig. 1: the physical system (i.e., the gas) being measured, denoted by A; the controller's (i.e., the demon's) internal state M (where the letter "M" stands for "Maxwell", "measurement apparatus" or "memory"); a classical register K recording the measurement's outcomes; and two independent baths B_1 and B_2 (one used during the feedback control stage, the other used for the final erasure of the measurement), which are assumed to be at the same finite temperature. This means that the overall process is assumed to be isothermal.

Without any feedback control, for isothermal processes, the second law of thermodynamics is equivalent to the statement that the work extracted from the system A can reach but not exceed the change in the free energy(These and other key concepts will be rigorously introduced and discussed in what follows. The purpose of these first few paragraphs is simply to provide a relatively informal overview of our main findings). of the system—in formula, $W^A_{\rm ext} \leqslant -\Delta F^A$. The main contribution of ref. 11 was to show that if feedback control is allowed instead, the work extracted can go all the way up to $W^A_{\rm ext} = -\Delta F^A + \beta^{-1} I_{\rm QC}$, where $I_{\rm QC}$ is a non-negative term quantifying the amount of information collected by the measurement used to guide the subsequent feedback control protocol. In this sense, Maxwell's demon *can* indeed violate the second law of thermodynamics, but this conclusion should come as no surprise, since the demon is not yet included in the global thermodynamic balance at this point.

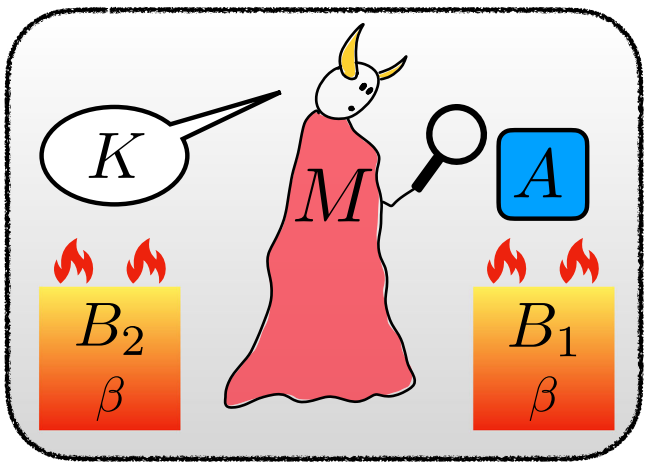

Fig. 1 | The systems appearing in our setup. the target system A, the controller (demon) consisting of an internal state M and a classical register K, and two baths B_1 and B_2 at the same inverse temperature β.

Indeed, once the demon itself is embodied in a physical system, such a violation of the second law turns out to be only a *local* violation, which is perfectly possible as long as it is compensated for elsewhere. According to Landauer's principle, such a compensation should be identified with the cost of performing the measurement and resetting the measurement apparatus and register at the end of the protocol, so that they are ready for use in the next round. Following this narrative, refs. 12,13 list a number of assumptions about the quantum feedback protocol so that, as one would expect, the work cost of implementing the measurement and performing its erasure is lower bounded as $W^{MK}_{\rm in} \geqslant \beta^{-1} I_{\rm QC}$, thus guaranteeing that the total net work extracted $W_{\rm tot} := W^A_{\rm ext} - W^{MK}_{\rm in} \leqslant -\Delta F^A$ is still within the limits of the second law of thermodynamics.

Our analysis begins by removing all assumptions from the consistency argument above. We argue that this is not just for the sake of mathematical generality, but is *necessary* for two reasons. The first reason is that, specifically in relation to refs. 11–13, some of the assumptions made therein are, as we will show in what follows, extremely restrictive—so stringent, in fact, that they are inconsistent in most cases, constraining the analysis to trivial situations. The second reason is a matter of principle: if certain assumptions are required to restore the validity of the second law, the consistency between thermodynamics and quantum information processing cannot be considered universal, contrary to what folklore claims.

We then show that, when all assumptions about the mathematical form of the quantum feedback protocol are removed, the work extracted from the target system is upper bounded as

$$W^A_{\rm ext} \leqslant -\Delta F^A_{0\to 4} + \beta^{-1} I_{\rm GO}\,, \tag{1}$$

while the work cost of implementing the measurement and its erasure is now lower bounded as

$$W^{MK}_{\rm in} \geqslant \beta^{-1}[\Delta S^{AMK} + I_{\rm GO}]\,, \tag{2}$$

where ΔS^{AMK} denotes the entropy change of the entire compound AMK due to the measurement process and $I_{\rm GO}$ is the *Groenewold–Ozawa information gain*[25,26]. Note that while the bound (1) looks similar to the one given in[11], the information quantity $I_{\rm GO}$ appearing in our bounds is different from the one used in refs. 11–13: in general, $I_{\rm GO} \lessgtr I_{\rm QC}$. But while $I_{\rm QC}$ does not provide the correct bounds in general, $I_{\rm GO}$ does and, moreover, gives the same numerical values as $I_{\rm QC}$ in all cases considered in[11–13]. Further, Eqs. (1) and (2) together imply that the net work extracted in general is bounded as

$$W_{\rm tot} := W^A_{\rm ext} - W^{MK}_{\rm in} \leqslant -\Delta F^A - \beta^{-1}\Delta S^{AMK}\,. \tag{3}$$

In other words, even if the final erasure is implemented in accordance with Landauer's principle, the second law may still be violated whenever $\Delta S^{AMK} < 0$. Eqs. (1) and (2) constitute the main technical contributions of this work: their formal statement is given as Theorem 1 below.

Finally, by means of explicit counterexamples, we show that the axioms of quantum theory *by themselves* are perfectly consistent with a measurement process that decreases the total entropy of the system-memory-register compound, implying a violation of the second law according to Eq. (3). This leads us to the main conceptual contribution of this work, i.e. the conclusion that—contrary to some cursory accounts—in a quantum mechanical feedback process it is not enough to eventually perform an erasure process, as stipulated by Landauer's principle, to guarantee the validity of the second law. In other words, *the second law of thermodynamics is logically independent of the axioms of quantum theory*, and its role is to constrain the set of possible measurement processes from the outset. Any attempt to *prove* the second law from *within* quantum theory is doomed to result in pure tautology[29,30].

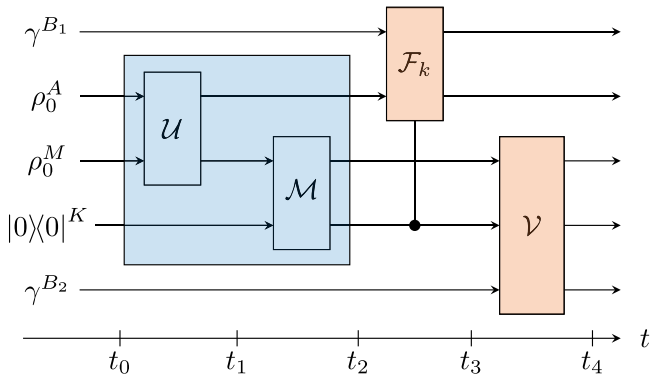

Fig. 2 | The circuit representation of a general quantum feedback control and erasure protocol. *Interaction step* ($t_0 \to t_1$): system A and memory M interact by a unitary channel $\mathcal{U}$. *Readout step* ($t_1 \to t_2$): an instrument $\mathcal{M}$ is applied on the memory M and the outcome k is written on the classical register K. The interaction step and the readout step together are referred to as the measurement step. *Feedback control step* ($t_2 \to t_3$): a controlled unitary channel $\mathcal{F}_k$ is applied on the compound of system A and thermal bath B_1 depending on the outcome k. *Erasure step* ($t_3 \to t_4$): a unitary channel $\mathcal{V}$ is applied on the compound of MK and thermal bath B_2, so as to return the state of MK to its initial configuration. The total compound system is assumed to evolve adiabatically during the entire protocol, that is, no heat is exchanged with any outside source.

Framework

Consider a quantum system Y associated with a finite-dimensional Hilbert space $\mathcal{H}^Y$. The algebra of linear operators L^Y on $\mathcal{H}^Y$ will be denoted as $\mathcal{L}(\mathcal{H}^Y)$, $\mathbb{1}^Y$ and $\mathbb{O}^Y$ denoting the unit and null operators, respectively. States on Y are represented by unit-trace positive operators, i.e., $\rho^Y \geqslant \mathbb{O}^Y$, $\mathrm{Tr}[\rho^Y] = 1$. A *thermodynamic system* Y is defined as the tuple $(\rho^Y; H^Y; \beta)$, where H^Y is the Hamiltonian and $\beta := 1/k_B T > 0$ is the inverse temperature of an external thermal bath, with k_B Boltzmann's constant. Throughout, we shall only consider the case where the thermal bath has a constant temperature, and so for notational simplicity we will abbreviate the thermodynamic system as $(\rho^Y; H^Y)$. When the system is in thermal equilibrium, the *thermal state* or *Gibbs state* is defined as $\gamma^Y := e^{-\beta H^Y}/Z^Y$, where $Z^Y := \mathrm{Tr}[e^{-\beta H^Y}]$ is the partition function.

The generalized quantum feedback control and erasure protocols we shall consider will comprise of five discrete time steps t_i, $i = 0, 1, 2, 3, 4$. The total system is composed of a target system A, a controller consisting of a memory M and a classical register K, and two thermal baths B_1, B_2, both of which have the same inverse temperature $\beta > 0$, as depicted in Figure 1. For notational simplicity, we shall omit superscripts when denoting any quantity pertaining to the *entire* compound B_1AMKB_2, reserving their use only when discussing subsystems; for example, the state of subsystem AMK at time step t_i will be denoted as $\rho_i^{AMK} := \mathrm{Tr}_{B_1B_2}[\rho_i]$, etc. In particular, we shall assume that the Hamiltonian at time step t_i reads $H_i = H^{B_1} + H_i^A + H^{MK} + H^{B_2}$. That is, at each time step we assume that there are no interaction terms between the different subsystems, and only the Hamiltonian of the target system A may change. The protocol is represented schematically in Figure 2; below we shall describe each step in detail.

The preparation step. At the initial time $t = t_0$, the compound system is prepared in the state

$$\rho_0 := \gamma^{B_1} \otimes \rho_0^A \otimes \rho_0^M \otimes |0\rangle\langle 0|^K \otimes \gamma^{B_2} \tag{4}$$

where ρ_0^A and ρ_0^M are arbitrary states on A and M, respectively, while $|0\rangle^K$ represents the idle state of the classical register(Note that the memory considered in ref. 12 is described by a Hilbert space with a direct sum structure. Here we describe the degrees of freedom of the labels of the blocks and the internal states of the memory using different quantum systems. In the context of our paper, the two pictures are clearly equivalent.), and $\gamma^{B_1}, \gamma^{B_2}$ are the thermal states of the baths, with respect to the same inverse temperature β. Note that a common assumption is that the initial state of the memory ρ_0^M is thermal at the same inverse temperature β as the two baths: while such an assumption is very reasonable from a physical point of view, and in particular facilitates the discussion of the erasure step (see below), for the sake of generality we keep ρ_0^M arbitrary.

The measurement step. This step comprises an interaction step and a readout step. The *interaction* or *pre-measurement* step (from $t = t_0$ to $t = t_1$) represents the interaction between A and M, described by a unitary channel $\mathcal{U}(\cdot) := U(\cdot)U^\dagger$ acting in AM. The *readout* or *pointer objectification* step (from $t = t_1$ to $t = t_2$) is represented as a *CP-instrument*[31] acting in M, namely, a family $\mathcal{M} := \{\mathcal{M}_k : k \in \mathcal{K}\}$ of completely positive linear maps $\mathcal{M}_k : \mathcal{L}(\mathcal{H}^M) \to \mathcal{L}(\mathcal{H}^M)$, labeled by the measurement outcomes $k \in \mathcal{K}$, such that their sum $\mathcal{M}_\mathcal{K} := \sum_{k\in\mathcal{K}} \mathcal{M}_k$ is trace-preserving, i.e., a channel. The instrument $\mathcal{M}$ is associated with a unique positive operator-valued measure (POVM) $\mathsf{M} := \{\mathsf{M}_k : k \in \mathcal{K}\}$, with elements defined using the "Heisenberg picture" dual of $\mathcal{M}_k$ as $\mathsf{M}_k := \mathcal{M}_k^*(\mathbb{1}^M)$. Since the POVM M acts in the memory, it is referred to as the *pointer observable*. After M is measured by the instrument $\mathcal{M}$, the observed outcome k is recorded in the classical register. Such classical readouts are assumed to be all perfectly distinguishable, and thus are represented, following a common convention in quantum information theory[32], by orthogonal pure states $|k\rangle^K$.

Accordingly, at $t = t_2$ the state of the compound system reads

$$\rho_2 := \gamma^{B_1} \otimes \left(\sum_{k\in\mathcal{K}} (\mathrm{id}^A \otimes \mathcal{M}_k)\left[\mathcal{U}(\rho_0^A \otimes \rho_0^M)\right] \otimes |k\rangle\langle k|^K \right) \otimes \gamma^{B_2} =: \sum_{k\in\mathcal{K}} p_k \rho_{2,k}\,, \tag{5}$$

where id^A denotes the identity channel acting in A, and

$$\rho_{2,k} := \gamma^{B_1} \otimes \rho_{2,k}^{AM} \otimes |k\rangle\langle k|^K \otimes \gamma^{B_2}\,,$$

with

$$\rho_{2,k}^{AM} := \frac{(\mathrm{id}^A \otimes \mathcal{M}_k)\left[\mathcal{U}(\rho_0^A \otimes \rho_0^M)\right]}{p_k}$$

whenever the probability of obtaining outcome k satisfies

$$p_k := \mathrm{Tr}\left\{(\mathrm{id}^A \otimes \mathcal{M}_k)\left[\mathcal{U}(\rho_0^A \otimes \rho_0^M)\right]\right\} > 0\,,$$

otherwise $\rho_{2,k}^{AM}$ can be defined arbitrarily.

We note that a fixed tuple $(\mathcal{H}^M, \rho_0^M, \mathcal{U}, \mathcal{M})$ defines a *measurement process* or *measurement scheme* for an instrument $\mathcal{A} := \{\mathcal{A}_k : k \in \mathcal{K}\}$ acting in the target system A, with the operations reading

$$\mathcal{A}_k(\cdot) := \mathrm{Tr}_M\left\{(\mathrm{id}^A \otimes \mathcal{M}_k)\left[\mathcal{U}(\cdot \otimes \rho_0^M)\right]\right\} \equiv \mathrm{Tr}_M\left[\mathbb{1}^A \otimes \mathsf{M}_k\, \mathcal{U}(\cdot \otimes \rho_0^M)\right]. \tag{6}$$

In particular, we stress that an instrument on the target system $\mathcal{A} := \{\mathcal{A}_k : k \in \mathcal{K}\}$ can be realized by means of infinitely many different measurement processes. One of the results of this work will be to show that the laws of thermodynamics constrain the latter, not the former.

Remark. The formalism of CP-instruments provides the most general readout (i.e., pointer objectification) procedure allowed by quantum theory. While general instruments in the target system A have been considered before, all previous works have focused on a restricted class of instruments acting in the memory M, namely, Lüders instruments compatible with a projection-valued measure (PVM), also known as "ideal projective measurements"[12–22]. M is a PVM if the effects M_k are mutually orthogonal projections, and the operations of the corresponding M-compatible Lüders

instrument read $\mathcal{M}^L_k(\cdot) := \mathsf{M}_k(\cdot)\mathsf{M}_k$. As shown by Ozawa[31], every instrument acting in A admits a canonical measurement scheme, where ρ^M_0 is chosen to be pure and the pointer observable is chosen to be a PVM. But we stress that the pointer observable in a given measurement process need not be a PVM; and, even if it is, the instrument measuring it need not be of the Lüders form. In fact, it is well known that every observable M admits infinitely many M-compatible instruments.

The feedback control step. From $t = t_2$ to $t = t_3$, a feedback control protocol is performed. This is implemented by coupling the compound AK with the thermal bath B_1 by a unitary channel $\mathcal{F}(\cdot) := F(\cdot)F^\dagger$, defined by the unitary operator (Note that unitarity of F implicitly assumes that K is represented by a Hilbert space $\mathcal{H}^K$ of dimension equal to the number of measurement outcomes, i.e, $\dim(\mathcal{H}^K) = |\mathcal{K}|$.).

$$F := \sum_{k\in\mathcal{K}} F_k \otimes |k\rangle\langle k|^K .$$

Here, F_k are unitary operators on B_1A, which induce the unitary channel $\mathcal{F}_k(\cdot) := F_k(\cdot)F_k^\dagger$ conditional on the classical register having recorded outcome k. At time step $t = t_3$, the state of the compound reads

$$\rho_3 := (\mathcal{F}^{B_1AK} \otimes \mathrm{id}^{MB_2})(\rho_2) = \sum_{k\in\mathcal{K}} p_k \rho_{3,k} , \tag{7}$$

where

$$\rho_{3,k} = \rho^{B_1AM}_{3,k} \otimes |k\rangle\langle k|^K \otimes \gamma^{B_2} .$$

Here, $\rho^{B_1AM}_{3,k} = (\mathcal{F}^{B_1A}_k \otimes \mathrm{id}^M)(\gamma^{B_1} \otimes \rho^{AM}_{2,k})$. We shall say that the feedback process is *pure unitary* if we choose $F_k = \mathbb{1}^{B_1} \otimes F^A_k$, so that for each outcome the target system undergoes an isolated unitary evolution. In other words, a pure unitary feedback process does not involve the thermal bath. This is the case considered in, e.g., refs. 11–13. However, since Szilard[2] onward, the traditional formulation typically considers a feedback protocol that is done in contact with a thermal bath, as we do here.

The erasure step. Lastly, the erasure process from $t = t_3$ to $t = t_4$ is modeled by coupling MK with the thermal bath B_2 by a unitary channel $\mathcal{V}(\cdot) := V(\cdot)V^\dagger$. We naturally assume that $H^A_3 = H^A_4$, since the target system A remains dormant. At time step t_4, the state of the compound system will read

$$\rho_4 := \left(\mathrm{id}^{B_1A} \otimes \mathcal{V}^{MKB_2}\right)(\rho_3) , \tag{8}$$

such that, by definition of "erasure", $\rho^{MK}_4 = \rho^{MK}_0 = \rho^M_0 \otimes |0\rangle\langle 0|^K$. That is, the interaction between MK and the bath B_2 returns the local state of MK back to its initial configuration. Such a setting appears in the context of *Landauer's principle*[4,33,34]. If, in addition, it holds that $\rho^{AMK}_4 = \rho^A_4 \otimes \rho^M_0 \otimes |0\rangle\langle 0|^K$, i.e., if the correlations between A and MK are also erased, then we say that the erasure is *perfect*. Otherwise, we call the erasure *partial*. While in principle perfect erasure can always be achieved if a suitable bath is provided, it is a non-trivial problem to determine whether such a unitary erasure process always exists for a *given* bath. To alleviate this problem, we also consider here protocols that include *partial* erasure. As mentioned above when discussing the preparation step, a conceptually simpler situation occurs when the initial state of the memory is thermal at the same bath temperature, so that the erasure process can be intuitively understood as a thermalization process.

About injected and extracted work, and the assumption of overall adiabaticity

The internal energy of a thermodynamic system is $E(\rho^Y; H^Y) := \mathrm{Tr}[\rho^Y H^Y]$, and the non-equilibrium free energy[35,36] is $F(\rho^Y; H^Y) := E(\rho^Y; H^Y) - \beta^{-1}S(Y)_\rho$, where $S(Y)_\rho := -\mathrm{Tr}[\rho^Y \ln \rho^Y]$ is the von Neumann entropy[23]. When a thermodynamic system transforms from $t = t_i$ to $t = t_j$ as $(\rho^Y_i; H^Y_i) \mapsto (\rho^Y_j; H^Y_j)$, we denote the increase in internal energy E, nonequilibrium free energy F, and entropy S as follows:

$$\Delta x^Y_{i\to j} := x\left(\rho^Y_j; H^Y_j\right) - x\left(\rho^Y_i; H^Y_i\right) \quad (x = E, F, S) . \tag{9}$$

Definition 1. Consider a thermodynamic system which transforms as $(\rho^Y_i; H^Y_i) \mapsto (\rho^Y_j; H^Y_j)$. The transformation is defined as *adiabatic* if it does not involve an exchange of heat with an external bath. In such a case, by the first law of thermodynamics, the work injected into (resp., extracted from) the system is defined as the increase (resp., decrease) in internal energy, i.e.,

$$W^Y_{\mathrm{in}} \equiv -W^Y_{\mathrm{ext}} := \Delta E^Y_{i\to j} .$$

In our formalism, all thermal baths (i.e., the systems B_1 and B_2) are treated as *internal* and so there are no *external* baths with which heat is exchanged. Moreover, following a well-established convention dating back to Szilard[2] and von Neumann[23], and routinely adopted until these days[11,12,16,37–39], we assume that the pointer objectification implemented by the instrument $\mathcal{M}$ is also adiabatic, although it is obviously non-unitary. This may be justified if, for example, the objectification process is sufficiently fast with respect to the time scale required for heat to dissipate[40]. Concerning the rest, i.e., during the premeasurement, feedback, and erasure steps of the protocol, the total compound transforms by a global unitary channel which, by definition, does not involve an interaction with *any* external system, and so clearly no heat is exchanged here either. In conclusion, while the subsystem AMK exchanges heat with B_1 and B_2 during the feedback and erasure steps, respectively, we treat the total compound B_1AMKB_2 as transforming adiabatically during the entire protocol.

Since the total process is adiabatic, the net extracted work is identified with the decrease in internal energy of the entire compound, that is, $W_{\mathrm{tot}} = -\Delta E_{0\to 4}$. Now we wish to split the contribution to the total work as that originating from the target system A and that originating from the controller MK. To this end, we note that the target system is involved only during the measurement and feedback steps, the controller is involved only during the measurement and erasure steps, the thermal bath B_1 is involved only during the feedback step, and the thermal bath B_2 is involved only during the erasure step. As such, we may write (see Methods, Section IV A)

$$\begin{aligned} W_{\mathrm{tot}} &= -\Delta E_{0\to 4} \\ &= -\Delta E_{0\to 2} - \Delta E_{2\to 3} - \Delta E_{3\to 4} \\ &= -\Delta E^A_{0\to 2} - \Delta E^{MK}_{0\to 2} - \Delta E^{B_1A}_{2\to 3} - \Delta E^{MKB_2}_{3\to 4} \\ &= W^A_{\mathrm{ext}} - W^{MK}_{\mathrm{in}} , \end{aligned} \tag{10}$$

where

$$W^A_{\mathrm{ext}} := -\Delta E^A_{0\to 2} - \Delta E_{2\to 3} \equiv -\Delta E^A_{0\to 2} - \Delta E^{B_1A}_{2\to 3} \tag{11}$$

is the work extracted from the target system, and

$$W^{MK}_{\mathrm{in}} := \Delta E^{MK}_{0\to 2} + \Delta E_{3\to 4} \equiv \Delta E^{MK}_{0\to 2} + \Delta E^{MKB_2}_{3\to 4} \tag{12}$$

is the work injected into the controller.

General work bounds

Before providing general bounds for the work defined in Eqs. (11) and (12), let us first introduce some useful information-theoretic quantities. For any state ρ^A and a positive operator σ^A such that $\mathrm{m}supp(\rho^A) \subseteq \mathrm{m}supp(\sigma^A)$, the *Umegaki quantum relative entropy* is defined by $D(\rho^A \parallel \sigma^A) := \mathrm{Tr}[\rho^A(\ln \rho^A - \ln \sigma^A)] \geqslant 0$[41], which is non-negative due to Klein's inequality[42,43], and vanishes if and only if $\rho^A = \sigma^A$. The *quantum mutual information* of a bipartite state ρ^{AB} is defined as $I(A:B)_\rho := S(A)_\rho + S(B)_\rho - S(AB)_\rho \equiv D(\rho^{AB} \| \rho^A \otimes \rho^B) \geqslant 0$, with equality if and only if $\rho^{AB} = \rho^A \otimes$

ρ^B. On the other hand, the *conditional quantum entropy* of a bipartite state ρ^{AB} is defined as $S(A|B)_\rho := S(AB)_\rho - S(B)_\rho$, which can be negative. The *conditional quantum mutual information* of a tripartite state ρ^{ABC} is defined as $I(A:C|B)_\rho := S(A|B)_\rho + S(C|B)_\rho - S(AC|B)_\rho \geqslant 0$, where the non-negativity follows from the strong subadditivity of the von Neumann entropy (see, e.g., ref. 32). Finally, we introduce the following information measure related to the measurement process on the target system:

Definition 2. The *Groenewold–Ozawa information gain*[25,26] of the target system's measurement process is defined as:

$$I_{\mathrm{GO}} := S(A)_{\rho_0} - S(A|K)_{\rho_2}\,, \tag{13}$$

where the entropy of the post-measurement state of the target system conditioned by the classical register, $S(A|K)_{\rho_2}$, can equivalently be written as $\sum_k p_k S(\rho^A_{2,k})$, i.e., the average entropy of the posterior states of A.

Remark. Note that I_{GO} is determined entirely by the prior system state ρ^A_0 and the instrument $\mathcal{A}$ acting in A as defined in Eq. (6). The Groenewold–Ozawa information gain is guaranteed to be non-negative for all prior states ρ^A_0 if and only if the instrument $\mathcal{A}$ is *quasi-complete*; $\mathcal{A}$ is called quasi-complete if for all pure prior states ρ^A_0, the posterior states $\rho^A_{2,k} := \mathcal{A}_k(\rho^A_0)/p_k$ are also pure. An example of a quasi-complete instrument is an *efficient* instrument, whereby each operation can be written with a single Kraus operator, i.e., $\mathcal{A}_k(\cdot) = L_k(\cdot)L_k^\dagger$. In general, therefore, I_{GO} can be negative[26].

The following proposition gives universally valid expressions for the work associated with feedback control and erasure protocols with a general quantum measurement process, independent of thermodynamics and from a purely information-theoretic point of view.

Proposition 1. In the generalized quantum feedback control and erasure protocol (Fig. 2), the extracted work from the system is

$$W^A_{\mathrm{ext}} = -\Delta F^A_{0\to 4} + \beta^{-1}\left[I_{\mathrm{GO}} - I(A:K)_{\rho_3} - S^{B_1}_{\mathrm{irr}}\right], \tag{14}$$

and the work needed to run the controller is

$$W^{MK}_{\mathrm{in}} = \beta^{-1}\Delta S^{AMK}_{0\to 2} + \beta^{-1}\left[I_{\mathrm{GO}} + I(A:M|K)_{\rho_2} + S^{B_2}_{\mathrm{irr}}\right], \tag{15}$$

where

$$S^{B_1}_{\mathrm{irr}} := \sum_{k\in\mathcal{K}} p_k\left(I(A:B_1)_{\rho_{3,k}} + D\left(\rho^{B_1}_{3,k} \parallel \gamma^{B_1}\right)\right) \geqslant 0,$$
$$S^{B_2}_{\mathrm{irr}} := I(MK:B_2)_{\rho_4} + D(\rho^{B_2}_4 \parallel \gamma^{B_2}) \geqslant 0,$$

denote the irreversible entropy production associated with the isothermal feedback and erasure steps.

See Methods, Section IV B, for the proof. We immediately see that Eq. (14) contains, besides the usual free energy change, a correction term that arises from the specific implementation of measurement and feedback protocol. Similarly, Eq. (15) contains additional correction terms to the usual entropy change of target system and controller.

We note that an equality similar to Eq. (14) was obtained in ref. 14, except that there the entropy production $S^{B_1}_{\mathrm{irr}}$ as well as the mutual information $I(A:K)_{\rho_3}$ was missing. The term $I(A:K)_{\rho_3} = S(\rho^A_3) - \sum_{k\in\mathcal{K}} p_k S(\rho^A_{3,k})$ corresponds to the *Holevo information* of the conditional states of A after feedback[44], which is non-negative and vanishes if and only if $\rho^A_{3,k} = \rho^A_3$ for all k. Reference [45] also derives a similar equality, but it uses the QC-mutual information, and not the Groenewold–Ozawa information gain.

From Proposition 1, by discarding terms that are always either positive or negative,we obtain universally valid bounds for injected and extracted work in quantum feedback control and erasure protocols, as well as necessary and sufficient conditions for their saturation.

Theorem 1. In the generalized quantum feedback control and erasure protocol (Fig. 2), the work extracted from the target system is upper bounded as

$$W^A_{\mathrm{ext}} \leqslant -\Delta F^A_{0\to 4} + \beta^{-1} I_{\mathrm{GO}}\,, \tag{16}$$

where the equality holds if and only if $I(A:K)_{\rho_3} = S^{B_1}_{\mathrm{irr}} = 0$.The work cost to run the controller is lower bounded as

$$W^{MK}_{\mathrm{in}} \geqslant \beta^{-1}\left[\Delta S^{AMK}_{0\to 2} + I_{\mathrm{GO}}\right], \tag{17}$$

where the equality holds if and only if $I(A:M|K)_{\rho_2} = S^{B_2}_{\mathrm{irr}} = 0$.

Remark. Let us discuss, by means of examples, the conditions under which the bounds in the above theorem can be saturated. A necessary condition for the equality in Eq. (16) is for the entropy production during the feedback step, $S^{B_1}_{\mathrm{irr}}$, to vanish. This will trivially be achieved if the feedback process is chosen to be pure unitary, i.e., so that for each outcome the target system undergoes an isolated unitary evolution, as assumed in[11]. However, note that in general this alone will not guarantee the other necessary condition for the equality in Eq. (16), i.e., a vanishing Holevo information $I(A:K)_{\rho_3}$. Recall that this quantity vanishes if and only if $\rho^A_{3,k} = \rho^A_3$ for all k, which implies that $\rho^A_{3,k} = \rho^A_{3,k'}$ for all k, k'. But if the feedback process is pure unitary, then $\rho^A_{3,k} = F^A_k(\rho^A_{2,k})F^{A\dagger}_k$. Since unitary channels leave the von Neumann entropy invariant, and two states are identical only if their entropies are identical, it clearly follows that a necessary condition for a vanishing Holevo information given a pure unitary feedback process is for all the posterior states after measurement, $\rho^A_{2,k}$, to have the same entropy. While this can be achieved if, for example, the system undergoes a von Neumann measurement of a non-degenerate observable, for general measurement processes this is not the case. This is why in physically relevant situations, in order to saturate Eq. (16) a feedback process that exchanges entropy with a thermal bath is required, thus going beyond the paradigm of pure unitary feedback processes employed in[11].

Remark. Similarly as above, a necessary condition for the equality in Eq. (17) is for the entropy production during erasure, $S^{B_2}_{\mathrm{irr}}$, to vanish. The other necessary condition, however, is given by a vanishing conditional mutual information $I(A:M|K)_{\rho_2} = \sum_{k\in\mathcal{K}} p_k\, I(A:M)_{\rho_{2,k}}$. Clearly, such a quantity vanishes if and only if $\rho^{AM}_{2,k} = \rho^A_{2,k}\otimes\rho^M_{2,k}$. Given that $\rho^{AM}_{2,k} = \mathrm{id}^A\otimes\mathcal{M}_k[\mathcal{U}(\rho^A_0\otimes\rho^M_0)]/p_k$, a sufficient condition for $I(A:M|K)_{\rho_2}$ to vanish is if the instrument $\mathcal{M}$ is *nuclear* (also known as *measure-and-prepare*[46] or *Gordon-Louisell type*[47]). That is, if it holds that $\mathcal{M}_k(\cdot) = \mathrm{Tr}[\mathcal{M}_k(\cdot)]\varrho^M_k$ for all k, where $\{\varrho^M_k\}$ is a fixed family of states on M. It is clear that a nuclear instrument acting in M will destroy the correlations between A and M for each outcome k. Every POVM admits a nuclear instrument and, as shown in Corollary 1 of[48] (see also Theorem 2 of[49]), if the pointer observable measured by $\mathcal{M}$ is rank-1, i.e., if all the effects $\mathsf{M}_k = \mathcal{M}^*_k(\mathbb{1}^M)$ are proportional to a rank-1 projection, then $\mathcal{M}$ is necessarily nuclear. Consequently, by choosing a rank-1 pointer observable, we can guarantee that the term $I(A:M|K)_{\rho_2}$ vanishes.

Comparison between the second law of thermodynamics and the second law of ITh

Our analysis so far has been independent of ther modynamics, but henceforth we will explore the consequences derived by combining the results of Proposition 1 with the second law of thermodynamics. Before doing so, however, we introduce two types of second laws of thermodynamics in this section, and show how they are related.

According to ref. 36, when a thermodynamic system Y transforms as $(\rho^Y_i; H^Y_i) \mapsto (\rho^Y_j; H^Y_j)$ by an isothermal processes, i.e., a process involving thermal baths with the same temperature, the second law can be formulated

as the following inequality:

$$W_{\text{ext}}^{Y} \leqslant -\Delta F_{i\to j}^{Y}\,. \tag{18}$$

Notice that the nonequilibrium free energy change in the right-hand side can be replaced by the change in *equilibrium* free energy $F_{\text{eq}}(H^Y) := -\beta^{-1}\ln Z^Y \equiv F(\gamma^Y; H^Y)$ whenever the initial state of Y is assumed to be in thermal equilibrium—this is a consequence of the implication[36]

$$\rho_i^Y = \gamma^Y \;\Rightarrow\; -\Delta F_{i\to j}^{Y} \leqslant -\Delta F_{\text{eq},i\to j}^{Y}. \tag{19}$$

The above inequality will be useful when connecting our analysis to previous ones.

The feedback control and erasure protocol we consider consists of the subsystem AMK interacting with baths B_1 and B_2, which are assumed to be of the same temperature so that the total process is isothermal. This is to ensure that our analysis falls within the domain of applicability of the second law as formulated in Eq. (18). As such, the feedback control and erasure protocol is consistent with the second law of phenomenological thermodynamics, or the *overall* second law, when the net extracted work given in Eq. (10) and the change in free energy of the compound AMK obey the relation in Eq. (18), i.e.,

$$W_{\text{tot}} \leqslant -\Delta F_{0\to 4}^{AMK}\,. \tag{20}$$

We remark again that the above inequality embodies the second law of thermodynamics when considered from the beginning (time t_0) to the end (time t_4) of the protocol, regardless of what happens in the intermediate steps. On the other hand, the feedback control and erasure protocol is consistent with the second law of ITh, as formulated in[12], when the net extracted work given in Eq. (10) is bounded by the change in free energy of *the target system alone*, i.e.,

$$W_{\text{tot}} \leqslant -\Delta F_{0\to 4}^{A}\,. \tag{21}$$

Since the memory and register are erased, the free energy change $\Delta F_{0\to 4}^{MK}$ is zero. Naively, one would be led to expect that $\Delta F_{0\to 4}^{AMK} = \Delta F_{0\to 4}^{A}$ as a result, thus suggesting that Eqs. (20) and (21) are equivalent. However, as the following proposition (proved in Methods, Section IV C) shows, they coincide *if and only if* erasure is perfect.

Proposition 2. The generalized quantum feedback control and erasure protocol (Figure 2) is consistent with the overall second law of thermodynamics, i.e., Eq. (20), if and only if

$$\Delta S_{0\to 2}^{AMK} \geqslant I(A:MK)_{\rho_4} - I(A:M|K)_{\rho_2} - I(A:K)_{\rho_3} - S_{\text{irr}}^{B_1} - S_{\text{irr}}^{B_2}. \tag{22}$$

Instead, the protocol is consistent with the generalized second law of ITh, i.e., Eq. (21), if and only if

$$\Delta S_{0\to 2}^{AMK} \geqslant -I(A:M|K)_{\rho_2} - I(A:K)_{\rho_3} - S_{\text{irr}}^{B_1} - S_{\text{irr}}^{B_2}. \tag{23}$$

Since $I(A:MK)_{\rho_4} \geqslant 0$, Eq. (22) always implies Eq. (23) : consistency with the second law of thermodynamics implies consistency with the second law of ITh. The converse implication holds if and only if erasure is perfect, i.e., $\rho_4^{AMK} = \rho_4^A \otimes \rho_0^M \otimes |0\rangle\langle 0|^K$.

In summary: a feedback control and erasure protocol that is consistent with the second law of phenomenological thermodynamics is guaranteed to also be consistent with the second law of ITh. However, if erasure is partial so that $I(A:MK)_{\rho_4} > 0$, then it may be the case that the protocol is consistent with the second law of ITh, but violates the second law of thermodynamics proper, allowing for work extraction beyond the Clausius bound.

When is a quantum measurement process compatible with the second law?

Proposition 2 above provides necessary and sufficient conditions for a given feedback control and erasure protocol to be consistent with the second law —be it the overall second law, or the second law of ITh. But now recall that a feedback control and erasure protocol is implemented by *first* performing a measurement, and *subsequently* performing feedback and erasure. It follows that, in order for a particular measurement process itself to be consistent with the second law(s), then all possible feedback control and erasure protocols that utilize that same measurement process must be consistent with the second law(s). This leads us to the following definition:

Definition 3. A given quantum measurement process

$$\rho_0^A \otimes \rho_0^M \otimes |0\rangle\langle 0|^K \mapsto \sum_{k\in\mathcal{K}} \left(\text{id}^A \otimes \mathcal{M}_k\right)\left[\mathcal{U}(\rho_0^A \otimes \rho_0^M)\right] \otimes |k\rangle\langle k|^K$$

is *compatible with the overall second law of thermodynamic* whenever Eq. (22) holds for all possible subsequent isothermal feedback and erasure processes. Similarly, the measurement process is *compatible with the second law of ITh* whenever Eq. (23) holds for all possible subsequent feedback and erasure processes.

We shall begin from a sufficient condition for a given measurement process to be compatible with the second law(s). As explicitly shown in Methods, Section IV D, we observe that the right hand side of Eq. (22) in Proposition 2 is never strictly positive, allowing us to obtain the following:

Proposition 3. A measurement process that does not decrease the total entropy, i.e., such that $\Delta S_{0\to 2}^{AMK} \geqslant 0$, is guaranteed to be compatible with the overall second law and, hence, also with the second law of ITh. Moreover, a sufficient condition for $\Delta S_{0\to 2}^{AMK} \geqslant 0$ to hold is if the instrument $\mathcal{M}$ responsible for pointer objectification implements a bistochastic channel, i.e., a CP linear map that preserves both the trace and the unit.

A consequence of Proposition 3 is that a feedback control and erasure protocol may violate the second law(s) only if it includes a measurement process that decreases the total entropy. However, it does not follow that *any* measuring process that decreases the entropy will *always* violate the second law(s). To this end, we obtain the following necessary condition for a measurement process to be compatible with the second law(s), proven in Methods, Section IV E:

Theorem 2. The measurement process is compatible with the second law of ITh if and only if

$$\Delta S_{0\to 2}^{AMK} \geqslant -I(A:M|K)_{\rho_2}, \tag{24}$$

or, equivalently,

$$\mathcal{H}(\{p_k\}) \geqslant I_{\text{GO}} + J_{\text{GO}}, \tag{25}$$

where $\mathcal{H}(\{p_k\}) := -\sum_{k\in\mathcal{K}} p_k \ln p_k$ is the Shannon entropy of the measurement outcomes probability distribution, and $J_{\text{GO}} := S(M)_{\rho_0} - S(M|K)_{\rho_2}$ is the Groenewold–Ozawa information gain of the memory.

Moreover, the above inequalities are necessary conditions for the measurement process to be compatible with the overall second law.

Eq. (24) states that even if the measurement process decreases the entropy, as long as the target system and memory are left in a sufficiently correlated state, then all possible feedback and erasure processes built on it will still be consistent with the second law of ITh. Such a condition is equivalently reformulated in Eq. (25) as a *tradeoff* between the information gains of the target system and the memory: if a given measurement process is compatible with the second law of ITh, then the information gain of the target system and that of the memory cannot be both arbitrarily large at the same time, but their sum must remain below the Shannon entropy of the measurement outcomes distribution.

Note that, in Theorem 2, it is the entropy of the *compoundAMK* that matters, not the entropy of the system A alone, which may well decrease as a result of the action of the effective instrument $\{\mathcal{A}_k : k \in \mathcal{K}\}$ in Eq. (6). In other words, the second law puts a restriction *on how a particular instrument is realized on the compound*, not on the instrument itself.

Remark. As stated in Proposition 3, if the instrument responsible for pointer objectification implements a bistochastic channel, then the entropy of the compound AMK is guaranteed not to decrease[50], thereby ensuring compatibility with the second law. A paradigmatic example of an objectification process that satisfies this condition is given by the Lüders instrument. But for any pointer observable M acting in the memory, there are M-compatible instruments which do not implement a bistochastic channel—for example, a nuclear instrument which prepares the memory in the same pure state for all outcomes. Additionally, let us recall that such an instrument will always destroy the correlations between system and memory, so that $I(A:M|K)_{\rho_2} = 0$, whereby a decrease in entropy is sufficient for the violation of the second law for some feedback and erasure process. In Methods, Section IV F, we explicitly construct such a feedback control and erasure protocol so that $\Delta S^{AMK}_{0\to 2}$ is strictly negative, and which violates both the overall second law of thermodynamics, as well as the second law of ITh.

As a consequence of the above, we see that the choice of the measurement process, in particular, of the objectification process, while not affecting the dynamics of the target system alone—which depends only on the pointer observable M, not on the choice of $\mathcal{M}$ implementing it, see Eq. (6) —instead has a *non-trivial thermodynamic implication*, since the state change of the memory enters directly into the thermodynamic balance. In fact, the common assumption that the pointer objectification is implemented by a Lüders instrument[12–22] obscures the role that the bistochasticity of such instruments plays in ensuring consistency with the second law, leading to the erroneous conclusion that the laws of quantum theory alone are sufficient to ensure compatibility with the second law. Here instead we have shown that, in order to obtain a full understanding of how the pointer objectification relates to the second law, the instrument $\mathcal{M}$ *must be treated as arbitrary*, as we have done, *lest one obtain statements of limited scope*.

Discussion

Here, we compare the work inequalities presented in Theorem 1 with those previously obtained by Sagawa and Ueda[11–13]. According to[11], the achievable upper bound on the amount of work extracted by feedback control from the target system A, assumed to be initially in equilibrium, is

$$W^A_{\rm ext} \leqslant -\Delta F^A_{{\rm eq},0\to 4} + \beta^{-1} I_{\rm QC}\,, \tag{26}$$

where $I_{\rm QC}$ is a nonnegative quantity named the *QC-mutual information*[11]. This quantity, in some particular situations, can be interpreted as a measure of the information gained by the measurement performed by the controller on the target system. Thus Eq. (26) implies that the second law (18) for system A can be violated in a feedback control protocol by an amount that is directly proportional to the information that the controller is able to obtain about the target system. Then, in a subsequent paper[12], the same authors showed that the quantity $\beta^{-1} I_{\rm QC}$, under suitable assumptions, provides a tight lower bound on the work cost for measurement and erasure:

$$W^{MK}_{\rm meas} + W^{MK}_{\rm eras} \;\equiv\; W^{MK}_{\rm in} \geqslant \beta^{-1} I_{\rm QC}\,. \tag{27}$$

Recalling that $W_{\rm tot} = W^A_{\rm ext} - W^{MK}_{\rm in}$, one thus obtains

$$W_{\rm tot} \leqslant -\Delta F^A_{{\rm eq},0\to 4}\,, \tag{28}$$

which ref. 12 refers to as the second law of ITh.

However, in order to be valid, the analysis presented by Sagawa and Ueda in refs. 11–13 requires the following assumptions on the quantum feedback control and erasure protocol:

Assumption 1. (**A-1**)[12]: The pointer objectification must be implemented by a Lüders instrument $\mathcal{M}^L_k(\cdot) := \mathsf{M}_k(\cdot)\mathsf{M}_k$ compatible with a projection valued measure M acting in M. That is, for each measurement outcome k, it must hold that

$$\rho^{AM}_{2,k} = \frac{(\mathbb{1}^A \otimes \mathsf{M}_k)\mathcal{U}(\rho^A_0 \otimes \rho^M_0)(\mathbb{1}^A \otimes \mathsf{M}_k)}{p_k}\,.$$

Assumption 2. (**A-2**)[11–13]: The instrument acting in the target system A, i.e., $\mathcal{A}_k(\cdot) := {\rm Tr}_M\left\{({\rm id}^A \otimes \mathcal{M}_k)\left[\mathcal{U}(\cdot \otimes \rho^M_0)\right]\right\}$, must be *efficient*. That is, every operation $\mathcal{A}_k$ must be expressible with only one Kraus operator.

Assumption 3. (**A-3**)[11]: The target system A must be initially prepared in the Gibbs state, that is, $\rho^A_0 = \gamma^A$.

Assumption 4. (**A-4**)[12]: At time step $t = t_2$, the target system and memory must be in a product state for each outcome k, i.e., $\rho^{AM}_{2,k} = \rho^A_{2,k} \otimes \rho^M_{2,k}$.

Assumption 5. (**A-5**)[11]: The feedback process must be pure unitary. That is, for each outcome k it must hold that $\rho^A_{3,k} = F^A_k(\rho^A_{2,k})F^{A\dagger}_k$.

Assumption 6. (**A-6**)[12]: The memory's Hilbert space and Hamiltonian possess a direct sum structure, i.e., $\mathcal{H}^M = \bigoplus_{k=0}^N \mathcal{H}^{M_k}$ and $H^M = \bigoplus_{k=0}^N H^{M_k}$, where $N = |\mathcal{K}|$ is the number of measurement outcomes, and H^{M_k} are Hamiltonians on the sector $\mathcal{H}^{M_k}$. Denoting the Gibbs states for each sector $\mathcal{H}^{M_k}$ as γ^{M_k}, it must hold that: (i) the initial state of the memory satisfies $\rho^M_0 = \gamma^{M_0}$, and (ii) the conditional states of the memory *before* erasure are thermal in the respective sectors, i.e., $\rho^M_{3,k} = \gamma^{M_k}$.

Note that none of the above assumptions need be satisfied by a general measurement and feedback process like that we consider. In fact, they are generally incompatible, except in trivial cases, as we discuss in the following remark.

Remark. First, assumptions (A-1) and (A-4) are typically incompatible, since given a Lüders-type pointer objectification, the post-measurement states $\rho^{AM}_{2,k}$ will in general be correlated. There are two cases in which (A-4) will be guaranteed to hold given (A-1): (i) if M_k are rank-1 projections, which is both necessary and sufficient for the M-compatible Lüders instrument $\mathcal{M}^L$ to be nuclear, then measurement of M by $\mathcal{M}^L$ is guaranteed to destroy the correlations between A and M; (ii) if the premeasurement unitary channel is local, i.e., $\mathcal{U} = \mathcal{U}^A \otimes \mathcal{U}^M$, then it trivially holds that $\rho^{AM}_{2,k} = p_k^{-1}\mathcal{U}^A(\rho^A_0) \otimes \mathsf{M}_k\mathcal{U}^M(\rho^M_0)\mathsf{M}_k$. But in such a case the measurement process does not extract any information at all, as it implements a trivial observable in A, namely, a POVM whose elements are all proportional to $\mathbb{1}^A$. Second, whenever the elements of the POVM measured by the instrument $\mathcal{A}$ in the target system are linearly independent (for example, if the observable is projection valued) then (A-1), (A-2), and (A-6) are compatible only if $\dim(\mathcal{H}^{M_0}) \leqslant N^{-1}\sum_{k=1}^N \dim(\mathcal{H}^{M_k})$. This follows from the fact that Gibbs states have full rank, and so the rank of $\rho^M_0 = \gamma^{M_0}$ equals $\dim(\mathcal{H}^{M_0})$, together with the fact that an efficient instrument compatible with an observable with linearly independent effects is *extremal*[51,52]. See Methods, Section IV G, for the proof. In particular, since M_k are projections onto the subspaces $\mathcal{H}^{M_k}$, then if M_k are rank-1 projections, which is necessary to guarantee compatibility of (A-1) and (A-4) discussed above, then $\mathcal{H}^{M_0}$ must also be 1-dimensional. In other words, in order to guarantee compatibility between assumptions (A-1), (A-2), (A-4), and (A-6), the initial state of the memory, ρ^M_0, must be pure. This is a physically unrealistic assumption due to the third law of thermodynamics[53].

On the other hand, as a consequence of our analysis, one easily sees that in fact Assumption (A-1) alone is already sufficient to obtain Eq. (21) which, under Assumption (A-3) and Eq. (19), directly implies Eq. (28). This is because Lüders channels are bistochastic, so that by Proposition 3 $\Delta S^{AMK}_{0\to 2} \geqslant 0$ is guaranteed to hold, which implies consistency with both second laws.

Thus, Eqs. (16) and (17) constitute a strict extension of Sagawa and Ueda's relations (26) and (27). This is because:

1. When the pointer objectification is implemented by a projective measurement on the memory, i.e., under (A-1), it holds that $\Delta S^{AMK}_{0\to 2} \geqslant 0$. Moreover, if $\Delta S^{AMK}_{0\to 2} > 0$, Eq. (17) is a more refined inequality than Eq. (27), and while the former can be saturated, the latter cannot.
2. When the instrument acting in A is assumed to be efficient, i.e., under (A-2), then the Groenewold–Ozawa information gain I_{GO} coincides with the QC-mutual information I_{QC}, as shown in ref. 54; in all other situations, the two quantities are unrelated, i.e., $I_{\mathrm{GO}} \lesseqgtr I_{\mathrm{QC}}$, but the one that retains its role in thermodynamic relations is I_{GO}.
3. When the target system is initialized in a Gibbs state, i.e., under (A-3), then $-\Delta F^{A}_{0\to 4} \leqslant -\Delta F^{A}_{\mathrm{eq},0\to 4}$ because of Eq. (19).In particular, we conclude that the correct information measure that remains valid for general measurement processes is I_{GO}, not I_{QC}. Although I_{GO} has been considered also in some previous works[14–16], these still imposed assumption (A-1). Our analysis shows that I_{GO} is the right quantity to consider even when (A-1) is not satisfied.

Summarizing, in this paper, we have shown that the consistency between the second law of thermodynamics and information processing is not guaranteed by the laws of quantum theory *simpliciter*. Instead, the second law must be taken as a primitive principle which imposes constraints on the physically valid quantum information processing protocols. In order to precisely characterize such constraints, we formulated quantum feedback control and erasure protocols with general isothermal feedback and general measurement processes. In particular, we did not assume that the pointer objectification step of the measurement process is implemented by a Lüders instrument, as was done in previous studies. We then provided necessary and sufficient conditions for such protocols to be consistent with the second law (Proposition 2). More generally, we provided necessary and sufficient conditions for a given measurement process to be consistent with the second laws for all subsequent feedback control and erasure processes (Proposition 3 and Theorem 2). These results show that while the second law is necessarily obeyed if the pointer objectification process is bistochastic—as is the case for Lüders instruments—the second law can be violated if the pointer objectification decreases the entropy, which is permitted by quantum theory alone. In this very sense, then, quantum theory alone is not a guarantee of compatibility with the second law.

Along the way, we derived expressions for the work extracted by feedback control and the work required for measurement and erasure (Proposition 1 and Theorem 1) which, unlike those presented in previous studies[11–16,18–20], are universally valid in the sense that we did not impose any assumptions on the feedback process, the measurement (including the pointer readout), or the initial state of the system. Of course, our equations recover those presented in previous studies[11–13], but are able to do so with fewer assumptions. As our other main result, we then show that the generalized second law of ITh presented here is guaranteed to hold for any quantum feedback control and erasure protocol that is consistent with the second law of thermodynamics proper, and that the two laws become equivalent in the case of perfect erasure of the demon's memory (Proposition 2).

This resolves the problem of the scope of the second law of ITh, which was unclear from previous studies, but can now be considered a *universally valid law of physics*. That is to say, since the conjunction of the second law and the laws of quantum theory implies that the second law of ITh will hold by logical necessity, as long as the second law and quantum theory are regarded as universally valid laws of physics, then so too must the second law of ITh be. Our results also contribute to the debate regarding the operational interpretation of the Groenewold–Ozawa information gain, which has been generally considered problematic, especially in those situations where it takes negative values; we have seen that this quantifies the amount by which the extractable work by measurement-plus-feedback exceeds the reduction in free energy[14,19], for all possible measurement and feedback processes.

An interesting direction to follow will be to look for applications of our approach to other formulations of the second law such as fluctuation theorems[15,55–59]. In the same way, another possible line for future research is to bring our analysis to the one-shot case[60–62], possibly beyond quantum theory[24,63,64], and to introduce insights from the thermodynamic reverse bound[34], retrodiction[58,59,65,66] and the theory of approximate recoverability[67]. Finally, an interesting line of future investigation will be to see how the second law of ITh interplays with the first and third laws of thermodynamics: the first law demands that the interaction between system and memory of the measuring device must be constrained so as to conserve the total energy, whereby the Wigner–Araki–Yanase theorem will impose limitations on the measurements one may perform[68–74]. On the other hand, the third law will prohibit the memory from being initialized in a pure state, which has also been shown to impose fundamental constraints on measurements[75–77]. While we have seen that the second law alone imposes no constraints on the measurements we can make on the target system—any instrument acting in the target system allows for a bistochastic measurement process that does not reduce the total entropy of the compound—it may be the case that, in conjunction with the other laws of thermodynamics, further constraints must be imposed on the quantum measurements that can be performed.

Methods

Preliminaries

Here, we introduce some preliminary concepts which will be used in the technical proofs appearing throughout the rest of the manuscript.

Definition 4. Consider a thermodynamic system $(\rho^A; H^A)$. The *internal energy* is defined as

$$E(\rho^A; H^A) := \mathrm{Tr}\left[\rho^A H^A\right],$$

and the *nonequilibrium free energy*[35,36] is defined as

$$F(\rho^A; H^A) := E(\rho^A; H^A) - \beta^{-1} S(A)_\rho \equiv F_{\mathrm{eq}}(H^A) + \beta^{-1} D(\rho^A \| \gamma^A),$$

where $F_{\mathrm{eq}}(H^A) := -\beta^{-1} \ln Z^A \equiv F(\gamma^A; H^A)$ is the equilibrium (Helmholtz) free energy.

Lemma 1. Consider a bipartite thermodynamic system $(\rho^{AB}; H^{AB})$. Assume that the Hamiltonian is additive, i.e., $H^{AB} = H^A + H^B := H^A \otimes \mathbb{1}^B + \mathbb{1}^A \otimes H^B$. It holds that

$$E\left(\rho^{AB}; H^{AB}\right) = E\left(\rho^A; H^A\right) + E\left(\rho^B; H^B\right)$$

and

$$F\left(\rho^{AB}; H^{AB}\right) = F\left(\rho^A; H^A\right) + F\left(\rho^B; H^B\right) + \beta^{-1} I(A:B)_\rho .$$

Proof. Note that by the definition of the partial trace, it holds that $\mathrm{Tr}[\rho^{AB} L^A \otimes \mathbb{1}^B] = \mathrm{Tr}[\rho^A L^A]$ for all L^A and ρ^{AB}. The additivity of the internal energy follows trivially from the additivity of the Hamiltonian. Now note that $F(\rho^{AB}; H^{AB}) = E(\rho^{AB}; H^{AB}) - \beta^{-1} S(AB)_\rho$. Observing that $S(AB)_\rho = S(A)_\rho + S(B)_\rho - I(A:B)_\rho$ completes the proof. ▪

Operations provide the most general description for how a quantum system may transform. In the Schrödinger picture, an operation acting in a system A is defined as a completely positive (CP), trace non-increasing linear map $\Phi : \mathcal{L}(\mathcal{H}^A) \to \mathcal{L}(\mathcal{H}^A)$. We shall denote the consecutive application of operations Φ_1 followed by Φ_2 as $\Phi_2 \circ \Phi_1$. For each operation, there exists a Heisenberg picture dual Φ^*, defined by the trace duality $\mathrm{Tr}[\Phi^*(L^A)\rho^A] = \mathrm{Tr}[L^A \Phi(\rho^A)]$ for all ρ^A and L^A. Φ^* is a sub-unital CP linear map, i.e., $\Phi^*(\mathbb{1}^A) \leqslant \mathbb{1}^A$. Among the operations are *channels*, which preserve the trace, and if Φ is a channel, then Φ^* is unital, i.e., $\Phi^*(\mathbb{1}^A) = \mathbb{1}^A$. We shall denote the identity channel acting in A as id^A, which satisfies $\mathrm{id}^A(L^A) = L^A$ for all L^A. An operation acting in a composite system AB is *local* if it can be written as $\Phi = \Phi^A \otimes \Phi^B$, such that $\Phi(L^A \otimes L^B) = \Phi^A(L^A) \otimes \Phi^B(L^B)$ for all L^A and L^B. As

such, $\Phi^A \otimes \mathrm{id}^B$ is an operation that acts locally and non-trivially only in subsystem A.

Lemma 2. Consider a bipartite thermodynamic system which transforms as $(\rho_i^{AB}; H_i^{AB}) \mapsto (\rho_j^{AB}; H_j^{AB})$, such that $\rho_j^{AB} = \Phi^A \otimes \mathrm{id}^B(\rho_i^{AB})$, where Φ^A is a channel acting in A and id^B is the identity channel acting in B. The following hold:

i. $\rho_j^A = \Phi^A(\rho_i^A)$ and $\rho_j^B = \rho_i^B$.
ii. If $H_k^{AB} = H_k^A + H^B$ for $k = i, j$, then $\Delta E_{i\to j}^{AB} = \Delta E_{i\to j}^A = \mathrm{Tr}[\Phi^A(\rho_i^A)H_j^A] - \mathrm{Tr}[\rho_i^A H_i^A]$.

Proof.

i. : For all L^A and L^B, it holds that

$$\begin{aligned}\mathrm{Tr}[\rho_j^A L^A] &= \mathrm{Tr}\left[\Phi^A \otimes \mathrm{id}^B(\rho_i^{AB})(L^A \otimes \mathbb{1}^B)\right] = \mathrm{Tr}\left[\rho_i^{AB}\Phi^{A*} \otimes \mathrm{id}^B(L^A \otimes \mathbb{1}^B)\right] \\ &= \mathrm{Tr}\left[\rho_i^{AB}\Phi^{A*}(L^A) \otimes \mathbb{1}^B\right] = \mathrm{Tr}\left[\rho_i^A \Phi^{A*}(L^A)\right] = \mathrm{Tr}\left[\Phi^A(\rho_i^A)L^A\right], \\ \mathrm{Tr}\left[\rho_j^B L^B\right] &= \mathrm{Tr}\left[\Phi^A \otimes \mathrm{id}^B(\rho_i^{AB})(\mathbb{1}^A \otimes L^B)\right] = \mathrm{Tr}\left[\rho_i^{AB}\Phi^{A*} \otimes \mathrm{id}^B(\mathbb{1}^A \otimes L^B)\right] \\ &= \mathrm{Tr}\left[\rho_i^{AB}\mathbb{1}^A \otimes L^B\right] = \mathrm{Tr}\left[\rho_i^B L^B\right].\end{aligned}$$

Here, we have used the definition of the partial trace, the trace duality, and the fact that Φ^{A^*} is unital while $\mathrm{id}^B(L^B) = L^B$ for all L^B. Since $\mathrm{Tr}[\rho^A L^A] = \mathrm{Tr}[\sigma^A L^A]$ for all L^A if and only if $\rho^A = \sigma^A$ completes the proof.

ii. This follows from item (i), together with the additivity of the Hamiltonian, Lemma 1, and the fact that $H_i^B = H_j^B = H^B$.□

Lemma 3. Consider a system Y and a thermal bath B, which transform as $(\rho_i^{YB}; H_i^{YB}) \mapsto (\rho_j^{YB}; H_j^{YB})$. Assume that $\rho_i^{YB} := \rho_i^Y \otimes \gamma^B$, and that $\rho_j^{YB} = \Phi(\rho_i^{YB})$ with $\Phi(\cdot) := U(\cdot)U^\dagger$ a unitary channel, and that $H_k^{YB} := H_k^Y + H^B$ for $k = i, j$. Then the extracted work from system Y will read

$$W_{\mathrm{ext}}^Y = -\Delta E_{i\to j}^{YB} = -\Delta F_{i\to j}^Y - \beta^{-1} S_{\mathrm{irr}}^B,$$

where

$$S_{\mathrm{irr}}^B := I(Y:B)_{\rho_j} + D(\rho_j^B \parallel \gamma^B) \geqslant 0$$

is the irreversible entropy production, vanishing if and only if $\rho_j^{YB} = \rho_j^Y \otimes \gamma^B$.

Proof.. Since unitary evolution is adiabatic, then by Definition 1 the extracted work from the compound YB will equal the decrease in internal energy, and so by Definition 4 it holds that $W_{\mathrm{ext}}^{YB} := -\Delta E_{i\to j}^{YB} = -\Delta F_{i\to j}^{YB} - \beta^{-1}\Delta S_{i\to j}^{YB} = -\Delta F_{i\to j}^{YB}$, with the last step following from the fact that unitary evolution does not change the von Neumann entropy. Now note that by the first law of thermodynamics, it holds that $W_{\mathrm{ext}}^Y = -\Delta E_{i\to j}^Y - Q^Y$, where W_{ext}^Y is the work extracted from system Y, and $Q^Y := \Delta E_{i\to j}^B$ is the heat that flows to the bath B. By the additivity of the Hamiltonian and Lemma 1, it follows that $W_{\mathrm{ext}}^Y = -\Delta E_{i\to j}^Y - \Delta E_{i\to j}^B = -\Delta E_{i\to j}^{YB} =: W_{\mathrm{ext}}^{YB}$. We may therefore write

$$\begin{aligned}W_{\mathrm{ext}}^Y &= -\Delta F_{i\to j}^{YB} \\ &= -\Delta F_{i\to j}^Y - \Delta F_{i\to j}^B - \beta^{-1} I(Y:B)_{\rho_j} \\ &= -\Delta F_{i\to j}^Y - \beta^{-1}\left[I(Y:B)_{\rho_j} + D\left(\rho_j^B \parallel \gamma^B\right)\right].\end{aligned}$$

In the second line we have used Lemma 1 and the additivity of the Hamiltonian, together with the fact that system and bath are uncorrelated at initial time, and so $I(Y:B)_{\rho_i} = 0$. In the third line we use the fact that the bath is initially in thermal equilibrium, i.e., $\rho_i^B = \gamma^B$, together with Definition 4 and the fact that the bath Hamiltonian, and hence the bath equilibrium free energy, does not change. Finally, we recall that the mutual information $I(Y:B)_{\rho_j}$ is non-negative and vanishes if and only if $\rho_j^{YB} = \rho_j^Y \otimes \rho_j^B$, whereas the relative entropy $D(\rho_j^B \parallel \gamma^B)$ is non-negative and vanishes if and only if $\rho_j^B = \gamma^B$[43]. ▪

Proof of Proposition 1

We shall first prove Eq. (14). Given that feedback is implemented by a global unitary channel $\rho_2 \mapsto \rho_3 = \mathcal{F} \otimes \mathrm{id}^{MB_2}(\rho_2)$, the extracted work will read

$$\begin{aligned}W_{\mathrm{ext},2\to 3}^A &:= -\Delta E_{2\to 3} = \mathrm{Tr}[\rho_2 H_2] - \mathrm{Tr}[\mathcal{F} \otimes \mathrm{id}^{MB_2}(\rho_2)H_3] \\ &= \mathrm{Tr}\left[\rho_2^{B_1A}(H^{B_1} + H_2^A)\right] - \mathrm{Tr}\left[\rho_3^{B_1A}(H^{B_1} + H_3^A)\right] \\ &= \sum_{k\in\mathcal{K}} p_k\left(\mathrm{Tr}\left[\gamma^{B_1} \otimes \rho_{2,k}^A(H^{B_1} + H_2^A)\right] - \mathrm{Tr}\left[\mathcal{F}_k\left(\gamma^{B_1} \otimes \rho_{2,k}^A\right)(H^{B_1} + H_3^A)\right]\right) \\ &= -\sum_{k\in\mathcal{K}} p_k\left(\Delta F_{2\to 3,k}^A + \beta^{-1}\left[I(A:B_1)_{\rho_{3,k}} + D\left(\rho_{3,k}^{B_1} \parallel \gamma^{B_1}\right)\right]\right).\end{aligned} \tag{29}$$

Here, the second line follows from Lemma 2 and the fact that $\mathcal{F}$ acts locally in B_1AK, and that the Hamiltonian at t_2, t_3 is additive with only the Hamiltonian of A changing in time, and that the state of K does not change. The third line follows from Eq. (5) and Eq. (7). The final line follows from Lemma 3. Now let us note that we may write

$$\begin{aligned}-\sum_{k\in\mathcal{K}} p_k \Delta F_{2\to 3,k}^A &= \sum_{k\in\mathcal{K}} p_k\left(\mathrm{Tr}\left[\rho_{2,k}^A H_2^A\right] - \mathrm{Tr}\left[\rho_{3,k}^A H_3^A\right] + \beta^{-1}\left[S(\rho_{3,k}^A) - S(\rho_{2,k}^A)\right]\right) \\ &= \mathrm{Tr}\left[\rho_2^A H_2^A\right] - \mathrm{Tr}\left[\rho_3^A H_3^A\right] + \beta^{-1}\sum_{k\in\mathcal{K}} p_k\left[S\left(\rho_{3,k}^A\right) - S\left(\rho_{2,k}^A\right)\right] \\ &= \left(\mathrm{Tr}[\rho_2^A H_2^A] - \mathrm{Tr}\left[\rho_0^A H_0^A\right]\right) + \left(\mathrm{Tr}[\rho_0^A H_0^A] - \mathrm{Tr}\left[\rho_3^A H_3^A\right]\right) \\ &\quad + \beta^{-1}\left[I_{\mathrm{GO}} + S(\rho_3^A) - S(\rho_0^A) - I(A:K)_{\rho_3}\right] \\ &= \Delta E_{0\to 2}^A - \Delta F_{0\to 3}^A + \beta^{-1}\left[I_{\mathrm{GO}} - I(A:K)_{\rho_3}\right] \\ &= \Delta E_{0\to 2}^A - \Delta F_{0\to 4}^A + \beta^{-1}\left[I_{\mathrm{GO}} - I(A:K)_{\rho_3}\right].\end{aligned} \tag{30}$$

In the second line we use the fact that $\sum_{k\in\mathcal{K}} p_k \rho_{i,k}^A = \rho_i^A$. The third line is obtained by adding and subtracting $\mathrm{Tr}[\rho_0^A H_0^A]$, $\beta^{-1}S(\rho_0^A)$, and $\beta^{-1}S(\rho_3^A)$, and noting that $I_{\mathrm{GO}} = S(\rho_0^A) - \sum_{k\in\mathcal{K}} p_k S(\rho_{2,k}^A)$ and $I(A:K)_{\rho_3} = S(\rho_3^A) - \sum_{k\in\mathcal{K}} p_k S(\rho_{3,k}^A)$. The final line is obtained by noting that $\Delta F_{0\to 4}^A = \Delta F_{0\to 3}^A + \Delta F_{3\to 4}^A$, and that $\Delta F_{3\to 4}^A = 0$ since both the state and Hamiltonian of system A do not change between time step t_3 and t_4. Finally, since $W_{\mathrm{ext},0\to 2}^A = -\Delta E_{0\to 2}^A$, then by Eq. (29) and Eq. (30) we have that

$$\begin{aligned}W_{\mathrm{ext}}^A &= W_{\mathrm{ext},0\to 2}^A + W_{\mathrm{ext},2\to 3}^A \\ &= -\Delta F_{0\to 4}^A + \beta^{-1}\left(I_{\mathrm{GO}} - I(A:K)_{\rho_3} - \sum_{k\in\mathcal{K}} p_k\left[I(A:B_1)_{\rho_{3,k}} + D(\rho_{3,k}^{B_1} \parallel \gamma^{B_1})\right]\right),\end{aligned}$$

and so we obtain Eq. (14).

Next, we show Eq. (15). Since the erasure step is implemented by the global unitary channel $\rho_3 \mapsto \rho_4 = \mathrm{id}^{B_1A} \otimes \mathcal{V}(\rho_3)$, we have

$$\begin{aligned}W_{\mathrm{in},3\to 4}^{MK} &:= \Delta E_{3\to 4} = \mathrm{Tr}[\mathrm{id}^{B_1A} \otimes \mathcal{V}(\rho_3)H_4] - \mathrm{Tr}[\rho_3 H_3] \\ &= \mathrm{Tr}\left[\mathcal{V}\left(\rho_3^{MK} \otimes \gamma^{B_2}\right)(H^{MK} + H^{B_2})\right] - \mathrm{Tr}\left[\rho_3^{MK} \otimes \gamma^{B_2}(H^{MK} + H^{B_2})\right] \\ &= \Delta F_{3\to 4}^{MK} + \beta^{-1}\left[I(MK:B_2)_{\rho_4} + D(\rho_4^{B_2} \parallel \gamma^{B_2})\right] \\ &= -\Delta F_{0\to 2}^{MK} + \beta^{-1}\left[I(MK:B_2)_{\rho_4} + D(\rho_4^{B_2} \parallel \gamma^{B_2})\right].\end{aligned}$$

The second line follows from Lemma 2 and the fact that $\mathcal{V}$ acts locally in MKB_2, and the fact that the Hamiltonian at t_3, t_4 is additive while the Hamiltonians of MK and B_2 do not change. The third line follows from Lemma 3. The final line follows from the assumption of erasure, i.e., $\rho_4^{MK} = \rho_0^{MK}$, so that $\Delta F_{3\to 4}^{MK} = -\Delta F_{0\to 3}^{MK}$, together with the fact that both the state and Hamiltonian of MK do not change between time steps t_2 and t_3, so that $-\Delta F_{0\to 3}^{MK} = -\Delta F_{0\to 2}^{MK} - \Delta F_{2\to 3}^{MK} = -\Delta F_{0\to 2}^{MK}$. Given that

$W^{MK}_{\mathrm{in},0\to 2} = \Delta E^{MK}_{0\to 2} = \Delta F^{MK}_{0\to 2} + \beta^{-1}\Delta S^{MK}_{0\to 2}$, we have that

$$\begin{aligned} W^{MK}_{\mathrm{in}} &= W^{MK}_{\mathrm{in},0\to 2} + W^{MK}_{\mathrm{in},3\to 4} \\ &= \Delta F^{MK}_{0\to 2} + \beta^{-1}\Delta S^{MK}_{0\to 2} - \Delta F^{MK}_{0\to 2} + \beta^{-1}\left[I(MK:B_2)_{\rho_4} + D(\rho_4^{B_2} \parallel \gamma^{B_2})\right] \\ &= \beta^{-1}\left[\Delta S^{MK}_{0\to 2} + I(MK:B_2)_{\rho_4} + D(\rho_4^{B_2} \parallel \gamma^{B_2})\right]. \end{aligned} \tag{31}$$

Now note that in general, the following relationship holds:

$$\begin{aligned} I(A:M|K)_{\rho_2} &:= S(A|K)_{\rho_2} + S(M|K)_{\rho_2} - S(AM|K)_{\rho_2} \\ &= S(A|K)_{\rho_2} - S(A)_{\rho_0} + S(A)_{\rho_0} + S(MK)_{\rho_2} \\ &\quad - S(MK)_{\rho_0} + S(MK)_{\rho_0} - S(AMK)_{\rho_2} \\ &= S(A|K)_{\rho_2} - S(A)_{\rho_0} + S(MK)_{\rho_2} \\ &\quad - S(MK)_{\rho_0} + S(AMK)_{\rho_0} - S(AMK)_{\rho_2} \\ &= -I_{\mathrm{GO}} + \Delta S^{MK}_{0\to 2} - \Delta S^{AMK}_{0\to 2}\,. \end{aligned} \tag{32}$$

The second line is obtained by adding and subtracting $S(A)_{\rho_0}$ and $S(MK)_{\rho_0}$, together with the definition $S(AM|K)_{\rho_2} := S(AMK)_{\rho_2} - S(K)_{\rho_2}$ and $S(M|K)_{\rho_2} := S(MK)_{\rho_2} - S(K)_{\rho_2}$. The third line is obtained by noting the fact that $\rho_0^{AMK} = \rho_0^{A} \otimes \rho_0^{MK}$ so that $S(A)_{\rho_0} + S(MK)_{\rho_0} = S(AMK)_{\rho_0}$. By combining Eq. (32) and Eq. (31), we obtain the desired equality Eq. (15).

Proof of Proposition 2

By combining Eqs. (14) and (15), we obtain

$$\begin{aligned} W_{\mathrm{tot}} &= W^{A}_{\mathrm{ext}} - W^{MK}_{\mathrm{in}} \\ &= -\Delta F^{A}_{0\to 4} - \beta^{-1}\left(\Delta S^{AMK}_{0\to 2} + I(A:M|K)_{\rho_2} + I(A:K)_{\rho_3} + S^{B_1}_{\mathrm{irr}} + S^{B_2}_{\mathrm{irr}}\right). \end{aligned} \tag{33}$$

Recall that the protocol is consistent with the overall second law of thermodynamics if and only if $W_{\mathrm{tot}} \leqslant -\Delta F^{AMK}_{0\to 4}$. But now note that

$$\begin{aligned} -\Delta F^{AMK}_{0\to 4} &= -\Delta F^{A}_{0\to 4} - \Delta F^{MK}_{0\to 4} - \beta^{-1} I(A:MK)_{\rho_4} \\ &= -\Delta F^{A}_{0\to 4} - \beta^{-1} I(A:MK)_{\rho_4} \\ &\leqslant -\Delta F^{A}_{0\to 4}\,, \end{aligned} \tag{34}$$

where the first equality holds because of Lemma 1 and $I(A:MK)_{\rho_0} = 0$, the second equality follows from the erasure condition $\rho_4^{MK} = \rho_0^{MK}$, and the inequality follows from the non-negativity of the mutual information. Then by Eq. (33), the protocol is consistent with the overall second law if and only if

$$\Delta S^{AMK}_{0\to 2} + I(A:M|K)_{\rho_2} + I(A:K)_{\rho_3} + S^{B_1}_{\mathrm{irr}} + S^{B_2}_{\mathrm{irr}} \geqslant I(A:MK)_{\rho_4}. \tag{35}$$

By rearranging the above, we obtain Eq. (22). Now recall that the protocol is consistent with the second law of ITh if and only if $W_{\mathrm{tot}} \leqslant -\Delta F^{A}_{0\to 4}$. By the same arguments as before, only replacing $I(A:MK)_{\rho_4}$ in the right hand side of Eq. (35) with 0, we obtain Eq. (23).

It is clear that Eq. (22) and Eq. (23) are equivalent if and only if erasure is perfect, that is, $\rho_4^{AMK} = \rho_4^{A} \otimes \rho_0^{MK} = \rho_4^{A} \otimes \rho_0^{M} \otimes |0\rangle\langle 0|^{K}$, so that $I(A:MK)_{\rho_4} = 0$. But since in general $I(A:MK)_{\rho_4} \geqslant 0$, while Eq. (22) always implies Eq. (23), the converse implication does not always hold.

Proof of Proposition 3

To show that $\Delta S^{AMK}_{0\to 2} \geqslant 0$ is sufficient for compatibility of the measurement process with the overall second law, we must show that the right hand side of Eq. (22) is never strictly positive. Given the non-negativity of the irreversible entropy production terms $S^{B_1}_{\mathrm{irr}}, S^{B_2}_{\mathrm{irr}}$, it suffices to show that

$$I(A:MK)_{\rho_4} - I(A:M|K)_{\rho_2} - I(A:K)_{\rho_3} \leqslant 0.$$

To this end, let us note that

$$\begin{aligned} I(A:M|K)_{\rho_2} &= \sum_{k\in\mathcal{K}} p_k I(A:M)_{\rho_{k,2}} \\ &= \sum_{k\in\mathcal{K}} p_k D\left(\rho^{AM}_{2,k} \parallel \rho^{A}_{2,k} \otimes \rho^{M}_{2,k}\right) \\ &\geqslant \sum_{k\in\mathcal{K}} p_k D\left(\Lambda_k \otimes \mathrm{id}^{M}(\rho^{AM}_{2,k}) \parallel \Lambda_k \otimes \mathrm{id}^{M}(\rho^{A}_{2,k} \otimes \rho^{M}_{2,k})\right) \\ &= \sum_{k\in\mathcal{K}} p_k D\left(\rho^{AM}_{3,k} \parallel \rho^{A}_{3,k} \otimes \rho^{M}_{3,k}\right) \\ &= \sum_{k\in\mathcal{K}} p_k I(A:M)_{\rho_{k,3}} = I(A:M|K)_{\rho_3}. \end{aligned} \tag{36}$$

Here, $\Lambda_k(\cdot) := \mathrm{Tr}_{B_1}[\mathcal{F}_k(\gamma^{B_1} \otimes \cdot)]$ are the conditional channels acting in A during feedback, the third line follows from the data processing inequality[78], and the fourth line follows from item (i) of Lemma 2. Note that if feedback is pure unitary, so that $\Lambda_k(\cdot) = F^{A}_{k}(\cdot)F^{A\dagger}_{k}$, then the inequality above becomes an equality.

Now notice that the following equality holds from the chain rule:

$$I(A:M|K)_{\rho_3} + I(A:K)_{\rho_3} = I(A:MK)_{\rho_3}\,. \tag{37}$$

By Eq. (36) and Eq. (37), it follows that

$$\begin{aligned} I(A:MK)_{\rho_4} - I(A:M|K)_{\rho_2} - I(A:K)_{\rho_3} &\leqslant I(A:MK)_{\rho_4} - I(A:M|K)_{\rho_3} - I(A:K)_{\rho_3} \\ &= I(A:MK)_{\rho_4} - I(A:MK)_{\rho_3} \\ &= D(\rho_4^{AMK} \parallel \rho_4^{A} \otimes \rho_4^{MK}) - D(\rho_3^{AMK} \parallel \rho_3^{A} \otimes \rho_3^{MK}) \\ &= D(\mathrm{id}^{A} \otimes \Phi(\rho_3^{AMK}) \parallel \mathrm{id}^{A} \otimes \Phi(\rho_3^{A} \otimes \rho_3^{MK})) \\ &\quad - D(\rho_3^{AMK} \parallel \rho_3^{A} \otimes \rho_3^{MK}) \\ &\leqslant 0. \end{aligned}$$

Here, $\Phi(\cdot) := \mathrm{Tr}_{B_2}[\mathcal{V}(\cdot \otimes \gamma^{B_2})]$ is the erasure channel acting in MK, the fourth line follows from item (i) of Lemma 2, and the final line follows from the data processing inequality.

Now, recall from Proposition 2 that if a feedback control and erasure protocol is consistent with the overall second law, then it will necessarily also be consistent with the second law of ITh. Therefore, a measurement process satisfying $\Delta S^{AMK}_{0\to 2} \geqslant 0$ is guaranteed to be compatible with the second law of ITh.

Finally, we wish to show that if the instrument $\mathcal{M} := \{\mathcal{M}_k : k \in \mathcal{K}\}$ that is responsible for pointer objectification implements a bistochastic channel—a CP linear map that preserves both the trace and the unit—then $\Delta S^{AMK}_{0\to 2} \geqslant 0$ will necessarily hold. Note that the channel implemented by $\mathcal{M}$, i.e., $\mathcal{M}_{\mathcal{K}}(\cdot) := \sum_{k\in\mathcal{K}} \mathcal{M}_k(\cdot)$, is bistochastic if $\mathcal{M}_{\mathcal{K}}(\mathbb{1}^{M}) = \mathbb{1}^{M}$.

Recall that $\rho_2^{AMK} = \sum_{k\in\mathcal{K}} p_k\, \rho^{AM}_{2,k} \otimes |k\rangle\langle k|^{K}$. Since the classical register K is not entangled with AM, it follows that $S(K|AM)_{\rho_2} \geqslant 0$. Thus, we have

$$\begin{aligned} \Delta S^{AMK}_{0\to 2} &= S(AMK)_{\rho_2} - S(AMK)_{\rho_0} \\ &= S(AM)_{\rho_2} + S(K|AM)_{\rho_2} - S(AM)_{\rho_0} \\ &\geqslant S(AM)_{\rho_2} - S(AM)_{\rho_0}. \end{aligned}$$

Given that unitary channels are bistochastic, then so long as the channel $\mathcal{M}_{\mathcal{K}}$ is also bistochastic, then so too is the composition $\Theta := (\mathrm{id}^{A} \otimes \mathcal{M}_{\mathcal{K}})\mathcal{U}$. Now note that we may equivalently write the von Neumann entropy as $S(A)_{\rho} = -D(\rho^{A} \parallel \mathbb{1}^{A})$. As such, we have that

$$\begin{aligned} \Delta S^{AMK}_{0\to 2} &\geqslant S(AM)_{\rho_2} - S(AM)_{\rho_0} \\ &= D(\rho_0^{AM} \parallel \mathbb{1}^{AM}) - D(\rho_2^{AM} \parallel \mathbb{1}^{AM}) \\ &= D(\rho_0^{A} \otimes \rho_0^{M} \parallel \mathbb{1}^{AM}) - D(\Theta(\rho_0^{A} \otimes \rho_0^{M}) \parallel \mathbb{1}^{AM}) \\ &= D(\rho_0^{A} \otimes \rho_0^{M} \parallel \mathbb{1}^{AM}) - D(\Theta(\rho_0^{A} \otimes \rho_0^{M}) \parallel \Theta(\mathbb{1}^{AM})) \\ &\geqslant 0. \end{aligned}$$

Here, in the fourth line we have used the bistochasticity of Θ, and the final line follows from the data processing inequality.

A paradigmatic example of an objectification process that is bistochastic is given by the Lüders instrument. For any observable M, the operations of the corresponding M-compatible Lüders instrument read $\mathcal{M}^L_k(\cdot) := \sqrt{\mathsf{M}_k}(\cdot)\sqrt{\mathsf{M}_k}$. These are also known as "square-root measurements". It is clear that the channel implemented by a Lüders instrument is bistochastic, since $\mathcal{M}^L_{\mathcal{K}}(\mathbb{1}^M) = \sum_{k\in\mathcal{K}}\mathsf{M}_k = \mathbb{1}^M$. However, every observable admits instruments that are not of the Lüders type, but which nonetheless implement a bistochastic channel—for example, the instrument with operations $\mathcal{M}_k := \Phi\mathcal{M}^L_k$, where Φ is some arbitrary bistochastic channel.

Proof of Theorem 2

Let us note that the only term on the right hand side of Eq. (23) that is fixed by the measurement process alone is $-I(A:M|K)_{\rho_2}$. Therefore, the right hand side of this equation, given a fixed measurement process but for all possible subsequent feedback and erasure processes, is upper bounded as

$$-I(A:M|K)_{\rho_2} - I(A:K)_{\rho_3} - S^{B_1}_{\mathrm{irr}} - S^{B_2}_{\mathrm{irr}} \leqslant -I(A:M|K)_{\rho_2}. \tag{38}$$

This follows from the non-negativity of the mutual information and the entropy production terms. Imposing that the inequality in Eq. (23) must be satisfied even when the right hand side obtains the upper bound above, we thus arrive at Eq. (24). Note that the upper bound of Eq. (38) is achievable in the limit where feedback and erasure are quasistatic, so that $S^{B_1}_{\mathrm{irr}} = S^{B_2}_{\mathrm{irr}} = 0$, and such that for all measurement outcomes the feedback process transforms the target system to the same final state, i.e., $\rho^A_{3,k} = \rho^A_3$ for all k, so that $I(A:K)_{\rho_3} = 0$.

To show that Eq. (24) is equivalent to Eq. (25), let us note that

$$\begin{aligned}\Delta S^{AMK}_{0\to 2} &:= S(AMK)_{\rho_2} - S(AMK)_{\rho_0}\\ &= S(AM|K)_{\rho_2} + S(K)_{\rho_2} - S(AMK)_{\rho_0}\\ &= S(AM|K)_{\rho_2} + \mathscr{H}(\{p_k\}) - S(A)_{\rho_0} - S(M)_{\rho_0}.\end{aligned} \tag{39}$$

In the second line we use the definition of the conditional entropy, whereas in the final line we use the fact that $\rho^K_2 = \sum_k p_k|k\rangle\langle k|^K$ and that $\rho^{AMK}_0 = \rho^A_0 \otimes \rho^M_0 \otimes |0\rangle\langle 0|^K$. Moreover, let us note that by the definition of the conditional mutual information, it holds that

$$-I(A:M|K)_{\rho_2} = -S(A|K)_{\rho_2} - S(M|K)_{\rho_2} + S(AM|K)_{\rho_2}. \tag{40}$$

By inserting Eqs. (39) and (40) in Eq. (24) gives us Eq. (25).

Finally, we shall show that Eq. (24) is also necessary for the compatiblity of the measurement process with the overall second law. To this end, let us consider a feedback control and erasure protocol, and assume that the measurement process violates Eq. (24), i.e., assume that $\Delta S^{AMK}_{0\to 2} < -I(A:M|K)_{\rho_2}$, but such that the protocol is consistent with the overall second law, i.e., Eq. (22). This gives us the inequality

$$\Delta S^{AMK}_{0\to 2} > \Delta S^{AMK}_{0\to 2} + I(A:MK)_{\rho_4} - I(A:K)_{\rho_3} - S^{B_1}_{\mathrm{irr}} - S^{B_2}_{\mathrm{irr}}.$$

Assume also that the feedback and erasure processes are ideal and quasistatic, so that $I(A:K)_{\rho_3} = S^{B_1}_{\mathrm{irr}} = S^{B_2}_{\mathrm{irr}} = 0$. In such a case the above inequality becomes

$$\Delta S^{AMK}_{0\to 2} > \Delta S^{AMK}_{0\to 2} + I(A:MK)_{\rho_4}.$$

But by the non-negativity of the mutual information, this inequality cannot be satisfied. As such, if a measurement process violates Eq. (24), then it will necessarily violate Eq. (22) for some feedback and erasure process. It follows that Eq. (24) is necessary for compatibility of the measurement process with the overall second law.

A measurement process that is incompatible with the second laws

Recall from Proposition 3 that a necessary condition for the incompatibility of the measurement process with the second laws is that pointer objectification must not be bistochastic. This is always possible; for example, a measure and prepare instrument $\mathcal{M}_k(\cdot) = \mathrm{Tr}[\mathcal{M}_k(\cdot)]|\psi\rangle\langle\psi|^M$, where $|\psi\rangle^M$ is a fixed, arbitrary pure state of M. It is trivial that $\mathcal{M}_{\mathcal{K}}$ is not bistochastic, since $\mathcal{M}_{\mathcal{K}}(\mathbb{1}^M) = \mathrm{Tr}[\mathbb{1}^M]|\psi\rangle\langle\psi|^M \neq \mathbb{1}^M$. We shall now use a measurement process utilizing just such a pointer objectification, demonstrating that it is incompatible with the second laws.

Let $(\mathcal{H}^M, \rho^M_0, \mathcal{U}, \mathcal{M})$ be a measurement process for a Lüders instrument $\mathcal{A}^L_k(\cdot) := \mathsf{A}_k(\cdot)\mathsf{A}_k$, compatible with a projection valued measure A, acting in the target system. Here, we choose $\mathcal{M}$ to be compatible with a projection valued measure M, and we choose ρ^M_0 to be a mixed state, albeit of sufficiently low rank so that our model is in accordance with Proposition 4. Recall that any instrument $\mathcal{M}$ that is compatible with the same POVM will realize the same instrument acting in the target system. Therefore, let us choose this instrument to be nuclear, with the operations $\mathcal{M}_k(\cdot) = \mathrm{Tr}[\mathsf{M}_k\cdot]|\psi\rangle\langle\psi|^M$, where $|\psi\rangle^M$ is a fixed, arbitrary pure state of M. Now let us choose one particular outcome $k = h$, and choose the input state of the target system so that it has support only in the eigenvalue-1 eigenspace of the effect A_h, i.e., $\mathsf{A}_k\rho^A_0 = \delta_{k,h}\rho^A_0$. In such a case, it will hold that $p_k = \delta_{k,h}$, and so $\mathscr{H}(\{p_k\}) := -\sum_{k\in\mathcal{K}} p_k \ln p_k = 0$. Moreover, we have that $\rho^A_{2,k} = \delta_{k,h}\rho^A_0$, so that $I_{\mathrm{GO}} = 0$. But, given the choice of instrument $\mathcal{M}$ acting in the memory, it holds that $\rho^M_{2,k} = \delta_{k,h}|\psi\rangle\langle\psi|^M$, so that $J_{\mathrm{GO}} = S(M)_{\rho_0} > 0$. Our protocol therefore gives the inequality

$$\mathscr{H}(\{p_k\}) < I_{\mathrm{GO}} + J_{\mathrm{GO}},$$

which contradicts Eq. (25) and so, by Theorem 2, violates the second law of ITh and the overall second law for some feedback and erasure processes. Indeed, note also that in this model, we have $\rho^{AMK}_2 = \rho^A_0 \otimes |\psi\rangle\langle\psi|^M \otimes |h\rangle\langle h|^K$, where the lack of correlations between A and M follows from the fact that $\mathcal{M}$ is a nuclear instrument. In such a case, it holds that $\Delta S^{AMK}_{0\to 2} = -S(M)_{\rho_0} < 0$, whereas $-I(A:M|K)_{\rho_2} = -I(A:M)_{\rho_2} = 0$. It follows that

$$\Delta S^{AMK}_{0\to 2} < -I(A:M|K)_{\rho_2},$$

which contradicts Eq. (24).

Efficient instruments

Proposition 4. Let $(\mathcal{H}^M, \rho^M_0, \mathcal{U}, \mathcal{M})$ be a measurement scheme for an instrument $\mathcal{A}$ compatible with an observable $\mathsf{A} := \{\mathsf{A}_k : k = 0, \ldots, N\}$ acting in A, where N is the number of distinct measurement outcomes, and where $\mathsf{A}_0 = \mathbb{O}^A$ is a null effect. Assume that $\mathcal{M}$ is compatible with a projection valued measure $\mathsf{M} := \{\mathsf{M}_k : k = 0, \ldots, N\}$ acting in M, and denote $\mathcal{H}^{M_k} := \mathrm{supp}(\mathsf{M}_k)$. Assume that the effects of A, excluding the null effect A_0, are linearly independent. Then $\mathcal{A}$ is efficient only if

$$\mathrm{rank}\left(\rho^M_0\right) \leqslant \frac{\sum_{k=1}^N \dim(\mathcal{H}^{M_k})}{N} \leqslant \frac{\dim(\mathcal{H}^M)}{N},$$

with the second inequality becoming an equality if and only if $\mathsf{M}_0 = \mathbb{O}^M$.

Proof. Note that Assumption (A-6) assumes that the outcome associated with projecting M onto the subspace $\mathcal{H}^{M_0}$ is (statistically) never observed, i.e., it is observed with probability zero. For this reason, in what follows, we need to introduce the effect M_0 of the pointer observable, associated with a null effect $\mathsf{A}_0 = \mathbb{O}^A$ for the system observable, which makes the presentation a little cumbersome.

To prove the claim, we first note that an efficient instrument compatible with an observable with linearly independent effects is *extremal*[52]; given the instruments $\mathcal{A}, \mathcal{A}', \mathcal{A}''$, all with the same value space $\mathcal{K}$, $\mathcal{A}$ is extremal if

for any $\lambda \in (0, 1)$, we may write $\mathcal{A}_k(\cdot) = \lambda\mathcal{A}'(\cdot) + (1-\lambda)\mathcal{A}''(\cdot)$ only if $\mathcal{A} = \mathcal{A}' = \mathcal{A}''$. That is, an instrument $\mathcal{A}$ is extremal if it cannot be written as a convex combination of distinct instruments. As such, we shall first obtain necessary conditions on the rank of ρ_0^M that must be satisfied for the measurement scheme to implement a general extremal instrument $\mathcal{A}$.

Let us write $\rho_0^M = \sum_{i=1}^r q_i|\phi_i\rangle\langle\phi_i|$, where $|\phi_i\rangle$ are mutually orthogonal unit vectors, $\{q_i\}$ is a probability distribution, and $r = \mathrm{rank}(\rho_0^M)$. By linearity, for each i it holds that $(\mathcal{H}^M, |\phi_i\rangle, \mathcal{U}, \mathcal{M})$ is a measurement scheme for an instrument $\mathcal{A}^{(i)}$, such that $\sum_i q_i\mathcal{A}_k^{(i)}(\cdot) = \mathcal{A}_k(\cdot)$ for all k. Note that since outcome $k = 0$ of the pointer observable is associated with the null effect $\mathsf{A}_0 = \mathbb{O}^A$, then it holds that $\mathcal{A}_0(\cdot) = \mathcal{A}_0^{(i)}(\cdot) = \mathbb{O}^A$. Denoting the (projection) effects of the pointer observable M as $\mathsf{M}_k = \sum_\mu |\psi_{k,\mu}\rangle\langle\psi_{k,\mu}|$, where $\{|\psi_{k,\mu}\rangle\}$ is an orthonormal basis that spans $\mathcal{H}^M$, then for each i and k, by Eq. (6) we may write

$$\mathcal{A}_k^{(i)}(\cdot) = \mathrm{Tr}_M[(\mathbb{1}^A \otimes \mathsf{M}_k)U(\cdot \otimes |\phi_i\rangle\langle\phi_i|)U^\dagger] = \sum_\mu L_{k,\mu}^{(i)}(\cdot)L_{k,\mu}^{(i)\dagger},$$

where the Kraus operators read

$$L_{k,\mu}^{(i)} = V_{\psi_{k,\mu}}^\dagger U V_{\phi_i}.$$

Here, $V_\varphi : \mathcal{H}^A \to \mathcal{H}^A \otimes \mathcal{H}^M, |\xi\rangle \mapsto |\xi\rangle \otimes |\varphi\rangle$ are linear isometries defined by the unit vector $|\varphi\rangle \in \mathcal{H}^M$, which satisfy

$$V_\varphi^\dagger \mathbb{1}^{AM} V_{\varphi'} = \langle\varphi|\varphi'\rangle\mathbb{1}^A, \quad V_\varphi \mathbb{1}^A V_{\varphi'}^\dagger = \mathbb{1}^A \otimes |\varphi\rangle\langle\varphi'|.$$

Noting that $\sum_{k,\mu}|\psi_{k,\mu}\rangle\langle\psi_{k,\mu}| = \mathbb{1}^M$, it follows that for every $i \neq j$, it holds that

$$\sum_{k,\mu} L_{k,\mu}^{(i)\dagger}L_{k,\mu}^{(j)} = \sum_{k,\mu} V_{\phi_i}^\dagger U^\dagger V_{\psi_{k,\mu}}\mathbb{1}^A V_{\psi_{k,\mu}}^\dagger U V_{\phi_j} = V_{\phi_i}^\dagger \mathbb{1}^{AM} V_{\phi_j} = \mathbb{O}. \quad (41)$$

Let $\{L_{k,\nu} \mid \nu = 1, \ldots, R_k\}$ be a minimal Kraus representation for the operation $\mathcal{A}_k$, i.e., where $L_{k,\nu}$ are linearly independent and R_k is the Kraus-rank of $\mathcal{A}_k$. Note that since $\mathsf{A}_0 = \mathbb{O}^A$, then $L_{0,\nu} = \mathbb{O}^A$. Now assume that $\mathcal{A}$ is an extremal instrument. This implies that $\mathcal{A}_k = \mathcal{A}_k^{(i)}$ for all i and k. As shown in[79], for each i there exists an isometry $[u_{\mu,\nu}^{(i)} \in \mathbb{C}]$ such that

$$L_{k,\mu}^{(i)} = \sum_\nu u_{\mu,\nu}^{(i)} L_{k,\nu}, \quad \sum_\mu u_{\mu,\nu}^{(i)*} u_{\mu,\nu'}^{(i)} = \delta_{\nu,\nu'}. \quad (42)$$

By Eq. (41), Eq. (42), and orthonormality of $\{|\psi_{k,\mu}\rangle\}$, we may thus write for every $i \neq j$ the following:

$$\begin{aligned}\mathbb{O} &= \sum_{k,\mu,\mu'} L_{k,\mu}^{(i)\dagger}L_{k,\mu'}^{(j)}\langle\psi_{k,\mu}|\psi_{k,\mu'}\rangle \\ &= \sum_{k,\mu,\mu'}\left(\sum_\nu u_{\mu,\nu}^{(i)*}L_{k,\nu}^\dagger\right)\left(\sum_{\nu'} u_{\mu',\nu'}^{(j)}L_{k,\nu'}\right)\langle\psi_{k,\mu}|\psi_{k,\mu'}\rangle \quad (43)\\ &= \sum_{k,\nu,\nu'} L_{k,\nu}^\dagger L_{k,\nu'}\langle\psi_{k,\nu}^{(i)}|\psi_{k,\nu'}^{(j)}\rangle,\end{aligned}$$

where

$$|\psi_{k,\nu}^{(i)}\rangle := \sum_\mu u_{\mu,\nu}^{(i)}|\psi_{k,\mu}\rangle \in \mathrm{supp}(\mathsf{M}_k) \equiv \mathcal{H}^{M_k}. \quad (44)$$

As shown in[51], $\mathcal{A}$ is an extremal instrument if and only if the set

$$\{L_{k,\nu}^\dagger L_{k,\nu'} \mid k = 1, \ldots, N; \nu, \nu' = 1, \ldots, R_k\}$$

is linearly independent. As such, the equality condition in Eq. (43) holds only if $\langle\psi_{k,\nu}^{(i)}|\psi_{k,\nu'}^{(j)}\rangle = 0$ for all $k > 0$, ν, ν', and $i \neq j$. Now, by Eq. (42) and Eq. (44), together with the fact that $\langle\psi_{k,\mu}|\psi_{k',\mu'}\rangle = \delta_{k,k'}\delta_{\mu,\mu'}$, it is easily verified that $\langle\psi_{k,\nu}^{(i)}|\psi_{k',\nu'}^{(i)}\rangle = \delta_{k,k'}\delta_{\nu,\nu'}$ for every i. Indeed, since for every i, $|\psi_{k,\nu}^{(i)}\rangle \in \mathcal{H}^{M_k}$, then it also holds that $\langle\psi_{k,\nu}^{(i)}|\psi_{k',\nu'}^{(j)}\rangle = 0$ whenever $k \neq k'$. It follows that

$$\left\{|\psi_{k,\nu}^{(i)}\rangle \in \bigoplus_{k=1}^N \mathcal{H}^{M_k} \mid k = 1, \ldots, N; \nu = 1, \ldots, R_k; i = 1, \ldots, \mathrm{rank}(\rho_0^M)\right\}$$

must be a set of mutually orthogonal vectors. The cardinality of the above set is easily computed to be $\mathrm{rank}(\rho_0^M)\sum_{k=1}^N R_k$. But since $\bigoplus_{k=1}^N \mathcal{H}^{M_k}$ can only contain at most $\dim(\bigoplus_{k=1}^N \mathcal{H}^{M_k}) = \sum_{k=1}^N \dim(\mathcal{H}^{M_k})$ mutually orthogonal vectors, then $\mathcal{A}$ is extremal only if

$$\mathrm{rank}(\rho_0^M) \leqslant \frac{\sum_{k=1}^N \dim(\mathcal{H}^{M_k})}{\sum_{k=1}^N R_k}.$$

Now assume that $\mathcal{A}$ is an efficient instrument. It holds that $R_k = 1$ for each k, and $\mathcal{A}$ is an extremal instrument if and only if $\{L_k^\dagger L_k = \mathsf{A}_k \mid k = 1, \ldots, N\}$, i.e., the non-trivial effects of the measured observable A in A, are linearly independent. This completes the proof. □

Data availability

No datasets were generated or analyzed during the current study.

Received: 18 November 2023; Accepted: 26 November 2024;
Published online: 07 February 2025

References

1. Maxwell, J. C. *Theory of Heat* (Appleton, 1871).
2. Szilard, L. über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Z. Phys.* **53**, 840–856 (1929).
3. Brillouin, L. Maxwell's demon cannot operate: Information and entropy. i. *J. Appl. Phys.* **22**, 334–337 (1951).
4. Landauer, R. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **5**, 183–191 (1961).
5. Bennett, C. H. Logical reversibility of computation. *IBM J. Res. Dev.* **17**, 525–532 (1973).
6. Bennett, C. H. The thermodynamics of computation—a review. *Int. J. Theor. Phys.* **21**, 905–940 (1982).
7. Leff, H. & Rex, A. *Maxwell's Demon 2 Entropy, Classical and Quantum Information, Computing* (CRC Press, 2002) https://books.google.co.jp/books?id=ZE5uBwAAQBAJ.
8. Maruyama, K., Nori, F. & Vedral, V. Colloquium: the physics of Maxwell's demon and information. *Rev. Mod. Phys.* **81**, 1–23 (2009).
9. Ingarden, R. S., Kossakowski, A. & Ohya, M. *Information Dynamics and Open Systems* (Springer Dordrecht, 1997).
10. Sagawa, T. & Ueda, M. in *Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond* Ch. 6 (eds Klages, R., Just, W. & Jarzynski, C.) 181–211 (Wiley Online Library, 2013). https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527658701.ch6.
11. Sagawa, T. & Ueda, M. Second law of thermodynamics with discrete quantum feedback control. *Phys. Rev. Lett.* **100**, 080403 (2008).
12. Sagawa, T. & Ueda, M. Minimal energy cost for thermodynamic information processing: measurement and information erasure. *Phys. Rev. Lett.* **102**, 250602 (2009).
13. Sagawa, T. & Ueda, M. Erratum: minimal energy cost for thermodynamic information processing: measurement and information erasure [phys. rev. lett. 102, 250602 (2009)]. *Phys. Rev. Lett.* **106**, 189901 (2011).
14. Jacobs, K. Second law of thermodynamics and quantum feedback control: Maxwell's demon with weak measurements. *Phys. Rev. A* **80**, 012322 (2009).
15. Funo, K., Watanabe, Y. & Ueda, M. Integral quantum fluctuation theorems under measurement and feedback control. *Phys. Rev. E* **88**, 052121 (2013).

16. Abdelkhalek, K., Nakata, Y. & Reeb, D. Fundamental energy cost for quantum measurement. Preprint at https://arxiv.org/abs/1609.06981 (2016).
17. Strasberg, P., Schaller, G., Brandes, T. & Esposito, M. Quantum and information thermodynamics: a unifying framework based on repeated interactions. *Phys. Rev. X* **7**, 021003 (2017).
18. Mohammady, M. H. & Romito, A. Conditional work statistics of quantum measurements. *Quantum* **3**, 175 (2019).
19. Strasberg, P. Operational approach to quantum stochastic thermodynamics. *Phys. Rev. E* **100**, 022127 (2019).
20. Strasberg, P. Thermodynamics of quantum causal models: an inclusive, Hamiltonian approach. *Quantum* **4**, 240 (2020).
21. Strasberg, P. *Quantum Stochastic Thermodynamics: Foundations and Selected Applications* (Oxford University Press, 2022).
22. Latune, C. L. & Elouard, C. A thermodynamically consistent approach to the energy costs of quantum measurements. Preprint at http://arxiv.org/abs/2402.16037 (2024).
23. von Neumann, J. *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, 1955).
24. Minagawa, S., Arai, H. & Buscemi, F. Von neumann's information engine without the spectral theorem. *Phys. Rev. Res.* **4**, 033091 (2022).
25. Groenewold, H. J. A problem of information gain by quantal measurements. *Int. J. Theor. Phys.* **4**, 327–338 (1971).
26. Ozawa, M. On information gain by quantum measurements of continuous observables. *J. Math. Phys.* **27**, 759–763 (1986).
27. Danageozian, A., Wilde, M. M. & Buscemi, F. Thermodynamic constraints on quantum information gain and error correction: a triple trade-off. *PRX Quantum* **3**, 020318 (2022).
28. Mohammady, M. H. Thermodynamically free quantum measurements. *J. Phys. A: Math. Theor.* **55**, 505304 (2022).
29. Earman, J. & Norton, J. D. Exorcist XIV: the wrath of Maxwell's demon. Part I. From Maxwell to Szilard. *Stud. History Philosophy Sci. B: Stud. History Philosophy Mod. Phys.* **29**, 435–471 (1998).
30. Earman, J. & Norton, J. D. Exorcist XIV: the wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond. *Stud. History Philosophy Sci. B: Stud. History Philos. Mod. Phys.* **30**, 1–40 (1999).
31. Ozawa, M. Quantum measuring processes of continuous observables. *J. Math. Phys.* **25**, 79–87 (1984).
32. Wilde, M. M. *Quantum Information Theory* 2nd edn (Cambridge University Press, 2017).
33. Reeb, D. & Wolf, M. M. An improved Landauer principle with finite-size corrections. *New J. Phys* **16**, 103011 (2014).
34. Buscemi, F., Fujiwara, D., Mitsui, N. & Rotondo, M. Thermodynamic reverse bounds for general open quantum processes. *Phys. Rev. A* **102**, 032210 (2020).
35. Gaveau, B. & Schulman, L. A general framework for non-equilibrium phenomena: the master equation and its formal consequences. *Phys. Lett. A* **229**, 347–353 (1997).
36. Esposito, M. & Van den Broeck, C. Second law and Landauer principle far from equilibrium. *EPL (Europhys. Lett.)* **95**, 40004 (2011).
37. Mancino, L. et al. The entropic cost of quantum generalized measurements. *npj Quantum Inf.* **4**, 20 (2018).
38. Purves, T. & Short, A. J. Channels, measurements, and postselection in quantum thermodynamics. *Phys. Rev. E* **104**, 014111 (2021).
39. Panda, A., Binder, F. C. & Vinjanampathy, S. Nonideal measurement heat engines. *Phys. Rev. A* **108**, 062214 (2023).
40. Smith, A. et al. Verification of the quantum nonequilibrium work relation in the presence of decoherence. *New J. Phys.* **20**, 013008 (2018).
41. Umegaki, H. On information in operator algebras. *Proc. Japan Acad.* **37**, 459–461 (1961).
42. Klein, O. Zur quantenmechanischen begründung des zweiten hauptsatzes der wärmelehre. *Z. Phys.* **72**, 767–7775 (1931).
43. Nielsen, M. A. & Chuang, I. L.*Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).
44. Holevo, A. S. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems Inform. Transmission* **9**, 177–183 (1973).
45. Sagawa, T. Thermodynamics of information processing in small systems. *Prog. Theoret. Phys.* **127**, 1–56 (2012).
46. Horodecki, M., Shor, P. W. & Ruskai, M. B. Entanglement breaking channels. *Rev. Math. Phys.* **15**, 629–641 (2003).
47. Gordon, J. & Louisell, W. in *Physics of Quantum Electronics: Conference Proceedings* (eds Kelley, P., Lax, B. & Tannenwald, P.) 833–840 (McGraw-Hill, 1966).
48. Heinosaari, T. & Wolf, M. M. Nondisturbing quantum measurements. *J. Math. Phys.* **51**, 092201 (2010).
49. Pellonpää, J.-P. Quantum instruments: II. Measurement theory. *J. Phys. A Math. Theor.* **46**, 025303 (2013).
50. Alberti, P. & Uhlmann, A. *Stochasticity and Partial Order; Doubly Stochastic Maps and Unitary Mixing* (Springer Dordrecht, 1982).
51. D'Ariano, G. M., Perinotti, P. & Sedlák, M. Extremal quantum protocols. *J. Math. Phys.* **52**, 082202 (2011).
52. Pellonpää, J.-P. Quantum instruments: I. Extreme instruments. *J. Phys. A Math. Theor.* **46**, 025302 (2013).
53. Taranto, P. et al. Landauer Versus Nernst: what is the true cost of cooling a quantum system? *PRX Quantum* **4**, 010332 (2023).
54. Buscemi, F., Hayashi, M. & Horodecki, M. Global information balance in quantum measurements. *Phys. Rev. Lett.* **100**, 210504 (2008).
55. Sagawa, T. & Ueda, M. Fluctuation theorem with information exchange: role of correlations in stochastic thermodynamics. *Phys. Rev. Lett.* **109**, 180602 (2012).
56. Sagawa, T. & Ueda, M. Nonequilibrium thermodynamics of feedback control. *Phys. Rev. E* **85**, 021104 (2012).
57. Sagawa, T. & Ueda, M. Role of mutual information in entropy production under information exchanges. *New J. Phys.* **15**, 125012 (2013).
58. Buscemi, F. & Scarani, V. Fluctuation theorems from bayesian retrodiction. *Phys. Rev. E* **103**, 052111 (2021).
59. Aw, C. C., Buscemi, F. & Scarani, V. Fluctuation theorems with retrodiction rather than reverse processes. *AVS Quantum Sci.* **3**, 045601 (2021).
60. Horodecki, M. & Oppenheim, J. Fundamental limitations for quantum and nanoscale thermodynamics. *Nat. Commun.* **4**, 1–6 (2013).
61. Faist, P. & Renner, R. Fundamental work cost of quantum processes. *Phys. Rev. X* **8**, 021011 (2018).
62. Lipka-Bartosik, P. & Demkowicz-Dobrzański, R. Thermodynamic work cost of quantum estimation protocols. *J. Phys. A Math. Theor.* **51**, 474001 (2018).
63. Hänggi, E. & Wehner, S. A violation of the uncertainty principle implies a violation of the second law of thermodynamics. *Nat. Commun.* **4**, 1–5 (2013).
64. Krumm, M., Barnum, H., Barrett, J. & Müller, M. P. Thermodynamics and the structure of quantum theory. *New J. Phys.* **19**, 043025 (2017).
65. Buscemi, F., Schindler, J. & Šafránek, D. Observational entropy, coarse-grained states, and the Petz recovery map: information-theoretic properties and bounds. *New J. Phys.* **25**, 053002 (2023).
66. Bai, G., Šafránek, D., Schindler, J., Buscemi, F. & Scarani, V. Observational entropy with general quantum priors. *Quantum* **8**, 1524 (2024).
67. Buscemi, F., Das, S. & Wilde, M. M. Approximate reversibility in the context of entropy gain, information gain, and complete positivity. *Phys. Rev. A* **93**, 062314 (2016).
68. Wigner, E. P. Die messung quantenmechanischer operatoren. *Z. Physik A: Hadrons Nuclei* **133**, 101–108 (1952).
69. Araki, H. & Yanase, M. M. Measurement of quantum mechanical operators. *Phys. Rev.* **120**, 622–626 (1960).

70. Ozawa, M. Conservation laws, uncertainty relations, and quantum limits of measurements. *Phys. Rev. Lett.* **88**, 050402 (2002).
71. Tajima, H. & Nagaoka, H. Coherence-variance uncertainty relation and coherence cost for quantum measurement under conservation laws. Preprint at https://arxiv.org/abs/1909.02904 (2019).
72. Mohammady, M. H., Miyadera, T. & Loveridge, L. Measurement disturbance and conservation laws in quantum mechanics. *Quantum* **7**, 1033 (2023).
73. Kuramochi, Y. & Tajima, H. Wigner-Araki-Yanase Theorem for Continuous and Unbounded Conserved Observables. *Phys. Rev. Lett.* **131**, 210201 (2023).
74. Emori, H. & Tajima, H. Error and disturbance as irreversibility with applications: unified definition, Wigner–Araki–Yanase theorem and out-of-time-order correlator. Preprint at https://arxiv.org/abs/2309.14172 (2023).
75. Guryanova, Y., Friis, N. & Huber, M. Ideal projective measurements have infinite resource costs. *Quantum* **4**, 222 (2020).
76. Mohammady, M. H. & Miyadera, T. Quantum measurements constrained by the third law of thermodynamics. *Phys. Rev. A* **107**, 022406 (2023).
77. Mohammady, M. H. & Miyadera, T. Erratum: Quantum measurements constrained by the third law of thermodynamics [Phys. Rev. A 107, 022406 (2023)]. *Phys. Rev. A* **110**, 029901(E) (2024).
78. Schumacher, B. & Nielsen, M. A. Quantum data processing and error correction. *Phys. Rev. A* **54**, 2629 (1996).
79. Choi, M. D. Positive linear maps on complex matrices. *Linear Algebra Appl.* **10**, 285–290 (1975).

Acknowledgements

The authors would like to thank Arshag Danageozian, Marco Genoni, Masahito Hayashi, Kenta Koshihara, Yosuke Mitsuhashi, Nelly H.Y. Ng, Yoshifumi Nakata, Takahiro Sagawa, Valerio Scarani, Jeongrak Son, and Philipp Strasberg for their helpful comments and fruitful discussions. S.M. acknowledges the "Nagoya University Interdisciplinary Frontier Fellowship" supported by Nagoya University and JST, the establishment of university fellowships towards the creation of science technology innovation, Grant Number JPMJFS2120 and "THERS Make New Standards Program for the Next Generation Researchers" supported by JST SPRING, Grant Number JPMJSP2125. M.H.M. acknowledges support from the European Union under project ShoQC within ERA-NET Cofund in Quantum Technologies (QuantERA) program, from the Slovak Academy of Sciences under IMPULZ project No. IM-2023-79 (OPQUT), as well as from projects VEGA 2/0183/21 (DESCOM) and APVV-22-0570 (DeQHOST). K.K. acknowledges support from JSPS Grant-in-Aid for Early-Career Scientists, No. 22K13972; from MEXT-JSPS Grant-in-Aid for Transformative Research Areas (A) "Extreme Universe", No. 22H05254. F.B. acknowledges support from MEXT Quantum Leap Flagship Program (MEXT QLEAP) Grant No. JPMXS0120319794, from MEXT-JSPS Grant-in-Aid for Transformative Research Areas (A) "Extreme Universe" No. 21H05183, and from JSPS KAKENHI, Grants No. 20K03746 and No. 23K03230.

Author contributions

S.M. and M.H.M. contributed equally to the conception of the ideas and the execution of the work. K.S. contributed to the initial formulation of the ideas. K.K. contributed to clarifying the logic of the arguments presented. F.B. contributed to the conception of the ideas and the execution of the work. All authors contributed to the interpretation and discussion of the results.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Shintaro Minagawa, M. Hamed Mohammady, Kenta Sakai, Kohtaro Kato or Francesco Buscemi.