

Thermodynamic Constraints on Quantum Information Gain and Error Correction: A Triple Trade-Off

Arshag Danageozian^{1,*}, Mark M. Wilde¹, and Francesco Buscemi²

¹*Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, and Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana 70803, USA*

²*Department of Mathematical Informatics, Nagoya University, Furo-cho Chikusa-ku, Nagoya 464-8601, Japan*



(Received 14 December 2021; accepted 8 March 2022; published 26 April 2022)

Quantum error correction (QEC) is a procedure by which the quantum state of a system is protected against a known type of noise, by preemptively adding redundancy to that state. Such a procedure is commonly used in quantum computing when thermal noise is present. Interestingly, thermal noise has also been known to play a central role in quantum thermodynamics (QTD). This fact hints at the applicability of certain QTD statements in the QEC of thermal noise, which has been discussed previously in the context of Maxwell’s demon. In this paper, we view QEC as a quantum heat engine with a feedback controller (i.e., a demon). We derive an upper bound on the measurement heat dissipated during the error-identification stage in terms of the Groenewold information gain, thereby also providing the latter with a physical meaning when it is negative. Further, we derive the second law of thermodynamics in the context of this QEC engine, operating with general quantum measurements. Finally, we show that, under a set of physically motivated assumptions, this leads to a fundamental triple-trade-off relation, which implies a trade-off between the maximum achievable fidelity of QEC and the super-Carnot efficiency that heat engines with feedback controllers have been known to possess. A similar trade-off relation occurs for the thermodynamic efficiency of the QEC engine and the efficacy of the quantum measurement used for error identification.

DOI: [10.1103/PRXQuantum.3.020318](https://doi.org/10.1103/PRXQuantum.3.020318)

I. INTRODUCTION

The link between quantum thermodynamics (QTD) and information theory has been flourishing over many decades due to two important factors: the miniaturization of controllable systems, which has raised the importance of a detailed accounting of dissipated energy in quantum devices, and fruitful thoughts on an old problem known as Maxwell’s demon. The latter goes back to 1867, when Maxwell first suggested a challenge to the second law of thermodynamics, whereby an intelligent mechanism (i.e., a demon) is used to reduce the entropy of a gas without performing work on the system, seemingly violating the second law [1–3]. However, it was Szilard in 1929 who first made the connection between this entity and information theory. Assuming the correctness of the second law, Szilard showed that work extraction from a single

heat reservoir at temperature T , using Maxwell’s demon, requires the production of an amount of entropy equal to $k_B \ln 2$ per gas particle. In 1961, Landauer showed that the origin of this entropy production is due to the resetting of the memory of the demon, which requires dumping an amount of heat equal to $k_B T \ln 2$ into the thermal reservoir of temperature T . This has become the celebrated “Landauer erasure principle,” which has been extended to various scenarios in quantum information and quantum computing [4–9].

One interesting area in which Landauer’s principle has been applied is quantum error correction (QEC). The latter is of importance in quantum computation [10–12], where the state of the computational system (usually a collection of qubits) is corrected by first encoding it into a larger physical system (by adding redundant parts to the computational system) and then effectively reversing the action of the environmental noise using a recovery channel [13–17]. Hence entropic and thermodynamic analyses have been conducted [18–20] to ensure that the effective reverse dynamics in QEC does not violate the second law of thermodynamics. However, the question of whether the laws of thermodynamics can add upon the existing QEC literature has been an open problem.

*Corresponding author. arshag.danageozian@gmail.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

One major area where the field of thermodynamics has been contributing to QEC is in the theory of quantum measurements. In the latter, the “information” derived from a measurement can be quantified by the average entropy reduction of the measured system, which is also known as the Groenewold information gain [21,22]. Therefore, this quantity can serve as a measure of information gain during the error-detection stage of QEC. The Groenewold information gain also has a thermodynamic meaning in terms of dissipated measurement heat [23,24] but this has hitherto been contingent on it taking non-negative values.

A richer and more direct link between QTD and QEC has been known since the work of Ref. [25], where the latter can be understood as a heat engine with a feedback controller [26] (see Fig. 1). This link suggests that the laws of thermodynamics can be extended to the QEC setting. However, it is known that the second law for engines with feedback controllers varies slightly from that of typical Carnot engines. For instance, a critical thermodynamic feature of the former is that they can accomplish efficiencies higher than the Carnot efficiency [27]. Namely, Carnot’s formulation of the second law is modified for such engines, since they have the advantage of measurement and feedback that typical Carnot engines do not. Of course, such a modified statement about possible super-Carnot efficiencies has to be stated very carefully in order to avoid some of the historical inconsistencies mentioned above; this has been done with some success in previous literature [26,27]. However, one has to be certain to erase the memory of Maxwell’s demon (a classical register), as well as to systematically consider the measurement heat during the feedback process [6].

In this paper, we expand the link between QTD and QEC by deriving thermodynamic constraints on the tasks of information gain from a measurement and QEC. These results can be summarized as follows:

- (1) *Measurement heat is bounded from above by Groenewold information gain.* We show that, if the measurement apparatus is initialized in a thermal state (see, e.g., Ref. [28,29]), then the heat dissipated into it (from the measured system) during a general quantum measurement is bounded from above by the Groenewold information gain. This extends the existing thermodynamic interpretation of the Groenewold information gain to cases in which it is negative [22]. Previously, only efficient quantum measurements have been discussed, for which the Groenewold information gain is always non-negative and the usual interpretation applies [23,27,30].
- (2) *The second law of thermodynamics constrains quantum error correction.* We derive the second-law inequality in the context of QEC, which is accomplished in a more general setting than previously achieved in the literature [26,27]. More precisely, we allow for an arbitrary initial state of the system (which need not be a collection of qubits), arbitrary thermal noise [without an independent identically distributed (IID) assumption], and arbitrary generalized measurements at the error-detection stage. Achieving such a generalization is important to capture the full set of limitations that the laws of thermodynamics impose on QEC in a general physical setting.

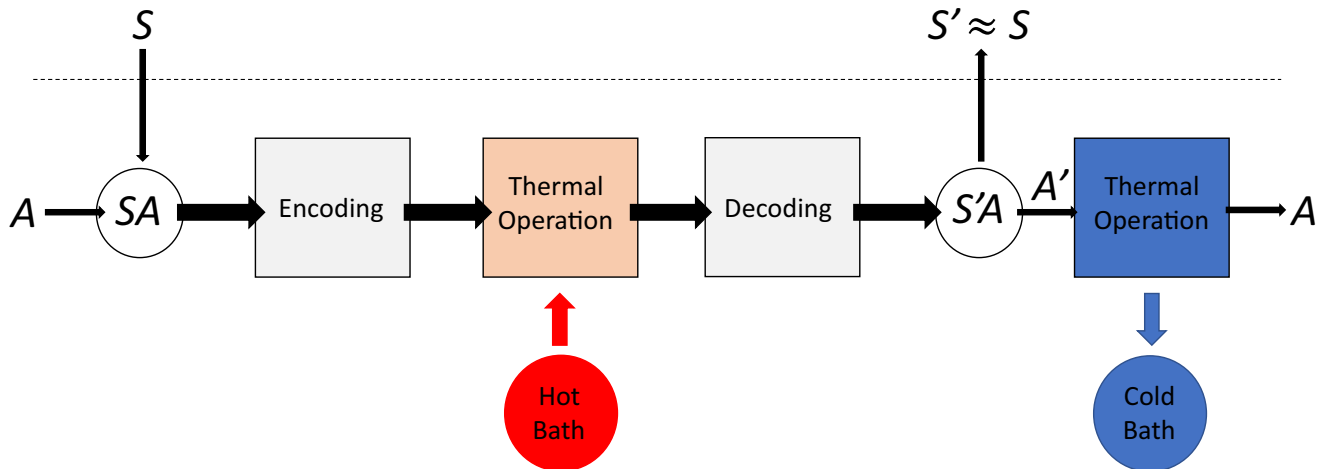


FIG. 1. A diagrammatic representation of QEC as a heat engine with a feedback controller, with the time flow going from left to right). From the left, we have the encoding of the system S using the ancillary system A for redundancy. Next, the encoded system SA interacts with a hot bath of temperature T_h (representing the noisy channel), followed by a decoding stage, where the errors in SA are detected and an attempt to correct them is made via a unitary feedback process. Finally, the ancillary system is returned (recycled) to its original state with the help of a cold bath of temperature T_c , in preparation for a future use.

(3) *Efficiency-fidelity trade-off in QEC engines.* We show that, under a set of physically motivated assumptions, the above thermodynamic limitations on QEC can be expressed in terms of a triple-trade-off relation between the maximum error-correction fidelity of a QEC engine (described by Fig. 1), the thermodynamic efficiency, and the efficacy of the quantum measurement at the error-detection stage of QEC. This can be broken up into two trade-off relations that can be summarized as follows: (i) the larger the thermodynamic efficiency beyond the Carnot limit, the smaller is the maximum achievable error-correcting fidelity of a QEC engine; and (ii) the larger the efficacy of the quantum measurement conducted at the error-detection stage, the smaller is the thermodynamic efficiency below the Carnot limit. As an additional result, we clarify the conditions under which one can arrive at a super-Carnot efficiency regime, improving upon the previous literature [27].

II. PRELIMINARIES

In what follows, we review various information-theoretic and thermodynamic quantities that are used in the rest of the paper.

Let $\mathcal{H}$ denote a Hilbert space, and let $\mathcal{L}(\mathcal{H})$ be the set of bounded linear operators acting on $\mathcal{H}$. Let $\mathcal{L}_+(\mathcal{H})$ denote the set of positive semidefinite operators acting on $\mathcal{H}$. The state of a physical system is described by a density matrix $\rho \in \mathcal{D}(\mathcal{H})$, where $\mathcal{D}(\mathcal{H})$ is the subset of linear operators $\mathcal{L}(\mathcal{H})$ that are positive semidefinite and have unit trace. We denote by $H(A)_\rho := -\text{Tr}[\rho^A \ln \rho^A]$ the von Neumann entropy of the density matrix $\rho^A \in \mathcal{D}(\mathcal{H}^A)$ of a system A . For a bipartite system AB described by a density matrix $\sigma^{AB} \in \mathcal{D}(\mathcal{H}^A \otimes \mathcal{H}^B)$, the quantum (von Neumann) conditional entropy and mutual information are denoted by $H(A|B)_\sigma := H(AB)_\sigma - H(B)_\sigma$ and $I(A:B)_\sigma := H(A)_\sigma - H(A|B)_\sigma$, respectively. Given two operators $M, N \in \mathcal{L}_+(\mathcal{H})$, we denote their quantum relative entropy by $D(M\|N)$, which is nonsymmetric and defined as

$$D(M\|N) := \text{Tr}[M \ln M] - \text{Tr}[M \ln N], \quad (1)$$

if $\text{supp}(M) \subseteq \text{supp}(N)$, and $+\infty$ otherwise, where $\text{supp}(\cdot)$ denotes the support (i.e., the orthogonal complement of the kernel) of an operator.

A linear map from $\mathcal{L}(\mathcal{H}^A)$ to $\mathcal{L}(\mathcal{H}^B)$ is denoted by $\mathcal{N}^{A \rightarrow B} : \mathcal{L}(\mathcal{H}^A) \rightarrow \mathcal{L}(\mathcal{H}^B)$. We say that a linear map is positive if $\mathcal{N}^{A \rightarrow B}(M^A) \in \mathcal{L}_+(\mathcal{H}^B)$ for all $M^A \in \mathcal{L}_+(\mathcal{H}^A)$, and trace preserving (TP) if $\text{Tr}[\mathcal{N}^{A \rightarrow B}(M^A)] = \text{Tr}[M^A]$ for all $M^A \in \mathcal{L}(\mathcal{H}^A)$. Similarly, we say that $\mathcal{N}^{A \rightarrow B}$ is trace nonincreasing (nondecreasing) if $\text{Tr}[\mathcal{N}^{A \rightarrow B}(M^A)] \leq (\geq) \text{Tr}[M^A]$ for all $M^A \in \mathcal{L}_+(\mathcal{H}^A)$. A linear map $\mathcal{N}^{A \rightarrow B}$

is unital if it maps the unit operator in $\mathcal{L}(\mathcal{H}^A)$ to the unit operator in $\mathcal{L}(\mathcal{H}^B)$, i.e., $\mathcal{N}^{A \rightarrow B}(I^A) = I^B$. Similarly, a linear map $\mathcal{N}^{A \rightarrow B}$ is subunital (superunital) if $\mathcal{N}^{A \rightarrow B}(I^A) \leq (\geq) I^B$. We define the adjoint map $(\mathcal{N}^{A \rightarrow B})^\dagger$ of any linear map $\mathcal{N}^{A \rightarrow B}$ through the relation $\text{Tr}[M \mathcal{N}^{A \rightarrow B}(N)] = \text{Tr}[(\mathcal{N}^{A \rightarrow B})^\dagger(M)N]$ for all $N \in \mathcal{L}(\mathcal{H}^A)$ and $M \in \mathcal{L}(\mathcal{H}^B)$. A linear map is unital if and only if the corresponding adjoint map is TP. Similarly, a linear map is subunital (superunital) if and only if the corresponding adjoint map is trace nonincreasing (nondecreasing). A positive linear map $\mathcal{N}^{A \rightarrow B}$ is called completely positive (CP) if for every Hilbert space $\mathcal{H}^R$, the map $\mathcal{I}^R \otimes \mathcal{N}^{A \rightarrow B}$ is positive, where $\mathcal{I}^R$ is the identity map acting on $\mathcal{L}(\mathcal{H}^R)$. It is well known that CP maps have a Kraus decomposition.

For any CP linear map $\mathcal{N}^{A \rightarrow B}$ acting on a density matrix $\rho^A \in \mathcal{D}(\mathcal{H}^A)$, we define the *efficacy* $\mathcal{E}$ [31–33] of the linear map $\mathcal{N}^{A \rightarrow B}$ with respect to ρ^A as

$$\mathcal{E}(\mathcal{N}^{A \rightarrow B}; \rho^A) := \text{Tr}[(\mathcal{N}^{A \rightarrow B})^\dagger \circ \mathcal{N}^{A \rightarrow B}(\rho^A)], \quad (2)$$

which is a measure of how “reversible” $\mathcal{N}^{A \rightarrow B}$ is with respect to ρ^A : indeed, it can be shown that the efficacy is related to the entropy change as follows [24,33]:

$$\begin{aligned} H(B)_{\mathcal{N}(\rho)} - H(A)_\rho &\geq D(\rho^A \| (\mathcal{N}^{A \rightarrow B})^\dagger \circ \mathcal{N}^{A \rightarrow B}(\rho^A)) \\ &= D(\rho^A \| \tilde{\rho}^A) - \ln \mathcal{E}(\mathcal{N}^{A \rightarrow B}; \rho^A) \\ &\geq -\ln \mathcal{E}(\mathcal{N}^{A \rightarrow B}; \rho^A), \end{aligned} \quad (3)$$

where

$$\tilde{\rho}^A := \frac{(\mathcal{N}^{A \rightarrow B})^\dagger \circ \mathcal{N}^{A \rightarrow B}(\rho^A)}{\text{Tr}[(\mathcal{N}^{A \rightarrow B})^\dagger \circ \mathcal{N}^{A \rightarrow B}(\rho^A)]} \in \mathcal{D}(\mathcal{H}^A), \quad (4)$$

is a valid density matrix. It is easy to see that, for efficacy to be strictly smaller (larger) than one for a fixed ρ^A , it is sufficient (but not necessary) for $\mathcal{N}^{A \rightarrow B}$ to be strictly subunital (superunital). Furthermore, due to Eq. (3), a subunital $\mathcal{N}^{A \rightarrow B}$ implies a positive entropy change but a superunital $\mathcal{N}^{A \rightarrow B}$ does not necessarily imply negative entropy change. See also Ref. [34] for a measure called diamond distance of nonunitarity, which is similar in spirit to the efficacy.

In our paper, we need a formalism that characterizes general quantum measurements, which can be thought of as a linear mapping between the pre- and postmeasurement states, along with some measurement statistics. For that, we define a *quantum instrument* [35–37] as a set $\{\mathcal{N}_x^A\}_{x \in \mathcal{X}}$ of completely positive trace-nonincreasing linear maps such that $\mathcal{N}^A = \sum_{x \in \mathcal{X}} \mathcal{N}_x^A$ is CPTP. Quantum instruments describe generalized quantum measurements with measurement outcomes $x \in \mathcal{X}$. The probability of each measurement outcome x is computed by $p_X(x) := \text{Tr}[\mathcal{N}_x^A(\rho^A)]$

and the corresponding postmeasurement state by

$$\rho^A \rightarrow \theta_x^A := \frac{\mathcal{N}_x^A(\rho^A)}{p_X(x)}. \quad (5)$$

It is convenient to record the measurement outcome in an auxiliary classical register X that is initially in a pure state $|0\rangle\langle 0|^X$. In this way, we can write the expected postmeasurement state as a quantum-classical state:

$$\theta^{AX} \equiv \mathcal{N}^{A \rightarrow AX}(\rho^A) = \sum_{x \in \mathcal{X}} \mathcal{N}_x^A(\rho^A) \otimes |x\rangle\langle x|^X, \quad (6)$$

where $\{|x\rangle^X\}_{x \in \mathcal{X}}$ is an orthogonal set of “pointer states” of the classical register.

Another important concept in this work is the notion of information gain of generalized quantum measurements. One such attempt in quantifying information gain from a measurement is given by the Groenewold information gain [21,22], which is defined as the expected entropy reduction of the system due to a general measurement described by the quantum instrument $\{\mathcal{N}_x^A\}_{x \in \mathcal{X}}$:

$$I_G(\{\mathcal{N}_x^A\}_x; \rho^A) := H(A)_\rho - \sum_{x \in \mathcal{X}} p_X(x) H(A)_{\theta_x} \quad (7)$$

$$= H(A)_\rho - H(A|X)_\theta. \quad (8)$$

When no confusion arises, in what follows we use the shorthand notation I_G instead of $I_G(\{\mathcal{N}_x^A\}_x; \rho^A)$. Although non-negative for projective measurements, the Groenewold information gain can, in general, become negative [22]. For example, this is easy to verify for an instrument that outputs the maximally mixed state $\theta_x^A = I^A/d^A$ for every measurement outcome $x \in \mathcal{X}$, independently of the input state. The negativity of the Groenewold information gain has arguably hindered its operational interpretation and, indeed, previous researchers have proposed alternative measures of information gain that are always non-negative, possess an operational interpretation, and reduce to Groenewold information gain for projective measurements. For example, Ref. [30] has defined the “quantum information gain” as the mutual information between the premeasurement purifying reference system R of A and the classical register X . It can be shown [30] that the Groenewold information gain I_G coincides with the quantum information gain (hence $I_G \geq 0$) for efficient measurements [i.e., when $\mathcal{N}_x(\cdot) = N_x(\cdot)N_x^\dagger$, which maps pure states to pure states]. Additionally, the quantum information gain has a compelling operational interpretation in terms of an information-processing task called measurement compression [38] (see also Refs. [39–43]).

As we end up discarding the classical register X after using the outcome to perform QEC, it is important to consider the entropy reduction due to this procedure, which is

given by

$$H(AX)_\theta - H(A)_\theta = H(X|A)_\theta \geq 0, \quad (9)$$

where the non-negativity follows from the fact that the quantum conditional entropy $H(X|A)_\theta = H(AX)_\theta - H(A)_\theta \geq 0$ is non-negative for every separable state. In this case, the state can be written as $\theta^{AX} = \sum_{x \in \mathcal{X}} p_X(x) \theta_x^A \otimes |x\rangle\langle x|^X$ and is thus separable. It is known that the quantity $H(X|A)$ has an information-theoretic meaning, as the compression rate in the task of classical data compression with quantum side information [44]. Thermodynamically, it is easy to see that the discarding process of X from the joint system AX , characterized by the entropy change $H(AX)_\theta - H(A)_\theta$, is equivalent to the erasure of the classical register state $\theta^X \rightarrow |0\rangle\langle 0|^X$ (or in other words, erasing the memory of Maxwell’s demon). For a study of thermodynamic erasure cost in the presence of a quantum memory, where quantum conditional entropy has been shown to be the optimal erasure cost, see Ref. [45].

In this paper, we also need to quantify the success of the QEC procedure: this can be accomplished by using a quantity called *entanglement fidelity* F_e [46,47]. If the initial state ρ_i^A of a system A is purified via a reference system R , i.e., $\rho_i^A \rightarrow \psi^{RA}$ with $\text{Tr}_R[\psi^{RA}] = \rho_i^A$, then the entanglement fidelity is defined as

$$F_e = \langle \psi^{RA} | \rho_f^{RA} | \psi^{RA} \rangle, \quad (10)$$

where $\rho_f^{RA} = [\text{id} \otimes (\mathcal{R} \circ \mathcal{E})](\psi^{RA})$ is the state of RA after the application of the noisy channel $\mathcal{E}$ and then the correcting channel $\mathcal{R}$. Physically, F_e is equal to the probability that the state ρ_f^{RA} passes a test for being the initial state ψ^{RA} (in particular, the test can be conducted by performing the measurement $\{|\psi\rangle\langle\psi|^{RA}, I^{RA} - |\psi\rangle\langle\psi|^{RA}\}$) [36]. Relevant properties of entanglement fidelity include being independent of the particular purification $|\psi\rangle^{RA}$ and being a lower bound on the input-output fidelity $F(\rho_f^A, \rho_i^A)$ [47], where the latter does not reflect the preservation (or the lack thereof) of entanglement that might be shared between A and some other system (e.g., R above).

It is important to note that when talking about QEC, it is not generally possible to correct for all possible preparations of the quantum state of A but, rather, for a subset $\rho^A \in \mathcal{C} \subseteq \mathcal{D}(\mathcal{H}^A)$ of such states, where $\mathcal{C}$ is known as the code space. Given a fixed noisy channel $\mathcal{E}$, a necessary and sufficient condition for the existence of a code space and a CPTP map $\mathcal{R}$ on that space that accomplishes perfect QEC ($F_e = 1$) is given by the Knill-Laflamme theorem [48,49]. It is also worth noting that the state of the reference system does not change as a result of any local CPTP operation $\mathcal{I}^R \otimes \mathcal{M}^A$ on A . This can be easily seen by checking that $\text{Tr}_A[(\mathcal{I}^R \otimes \mathcal{M}^A)(\psi^{RA})] = \text{Tr}_A[\psi^{RA}]$.

A major physical concept that plays a fundamental role in both fields of QTD and QEC is that of the irreversibility of the system-environment evolution $\rho_i^A \rightarrow \rho_f^A = \text{Tr}_B[U(\rho_i^A \otimes \sigma_i^B)U^\dagger]$, initiated from a product state. In QTD, this is typically quantified by $\Delta H(B)$, which is the net entropy dissipated into the environment B [5]. In QEC, this quantity has been given the name *entropy exchange* [46,47]. Here, the initial state of the environment is taken to be pure (without loss of generality [50]) and the initial state of A is purified using a reference system R . Then, the entropy exchange S_e is identified with the final state of RA , namely $S_e := H(RA)_{\rho_f}$. It can be shown that this quantity is independent of the purification and, more importantly, that it serves as an upper bound via $S_e \geq |\Delta H(B)|$ [47], where the inequality is saturated for a pure initial state of B . The latter property assigns a physical meaning to entropy exchange as the amount of entropy dissipated into an environment consisting of an initially pure state.

Finally, in order to show what constraints thermodynamics can add to QEC, we have to define the concepts of thermodynamic work and heat. A more detailed discussion is contained in Appendix A. Given two systems A and B , we can functionally separate a system of interest and a heat bath in the following way [51]: (i) we assume that the heat bath B is always prepared in an initially thermal state τ^B of temperature T ; and (ii) we assume that the system of interest A is controllable (in contrast to the heat bath), via an external control (work) parameter λ_t with predetermined trajectory $\{\lambda_t\}_{t \geq 0}$. The latter is reflected in the total Hamiltonian dynamics of AB by the fact that the local Hamiltonian of A (denoted by $\mathcal{H}^A$) depends implicitly on time via its explicit dependence on the work parameter λ_t , i.e., $\mathcal{H}_{\lambda_t}^A$. After making these two important distinctions between A and B , we define the internal energy $\mathcal{U}_t^A$ of A at time t , in the weak-coupling limit, as

$$\mathcal{U}_t^A = \text{Tr}[\rho_t^A \mathcal{H}_{\lambda_t}^A]. \quad (11)$$

From here, thermodynamic work and heat can be identified with two different contributions in the internal energy change of A [51] via

$$\partial_t \mathcal{U}_t^A = \text{Tr}[(\partial_t \rho_t^A) \mathcal{H}_{\lambda_t}^A] + \text{Tr}[\rho_t^A (\partial_t \mathcal{H}_{\lambda_t}^A)], \quad (12)$$

where the time integral of the first term defines heat Q_t^A and that of the second defines work W_t^A . To validate the intuition behind these definitions, note that if the system A is isolated (i.e., B is not present and hence we expect no heat to be dissipated), then ρ_t^A satisfies the Liouville–von Neumann equation and the first term in Eq. (12) becomes zero due to the cyclicity of trace $\text{Tr}[CD] = \text{Tr}[DC]$. On the other hand, if the external control parameter is not varied (hence we expect no work to be performed by or on A), then the second term in Eq. (12) becomes trivially equal to zero.

III. THERMODYNAMICS OF GENERAL QUANTUM INSTRUMENTS

It is known [35] that each CP map $\mathcal{N}_x^A$ of the quantum instrument $\{\mathcal{N}_x^A\}_{x \in \mathcal{X}}$ can be written in terms of a common unitary interaction between the system A and a measurement apparatus (probe) M , followed by a projective measurement on M (also known as an indirect measurement of system A [35,37,52]), as follows:

$$\mathcal{N}_x^A(\rho^A) = \text{Tr}_M[U^{AM}(\rho^A \otimes \sigma^M)(U^{AM})^\dagger P_x^M], \quad (13)$$

where $\sigma^M \in \mathcal{D}(\mathcal{H}^M)$ is the quantum state of the measurement apparatus, $\mathcal{H}^M$ is the Hilbert space of M , U^{AM} is a global unitary acting on the joint system AM , and $\{P_x^M\}_{x \in \mathcal{X}}$ is a complete set of projectors. We call the tuple $(\sigma^M, U^{AM}, \{P_x^M\}_{x \in \mathcal{X}})$ an indirect measurement model [35] (or implementation) of the quantum instrument $\{\mathcal{N}_x^A\}_{x \in \mathcal{X}}$. To a given quantum instrument, there is an infinite number of implementations, all yielding the same measurement statistics $\{p_X(x)\}_{x \in \mathcal{X}}$ and postmeasurement states $\{\theta_x^A\}_{x \in \mathcal{X}}$ of the system A but involving very different physical models in principle.

For the purposes of this paper, it is important to note that the quantum instrument formalism can also be thought of as a unitary interaction between the system A , a measurement apparatus M , and a classical register X , as

$$\theta^{AMX} = U^{AMX}(\rho^A \otimes \sigma^M \otimes |0\rangle\langle 0|^X)(U^{AMX})^\dagger, \quad (14)$$

with $U^{AMX} = V^{MX} U^{AM}$, where U^{AM} is the unitary operator in Eq. (13) and V^{MX} is another unitary operator performing a controlled transformation on the register system [6]

$$V^{MX} = \sum_{x \in \mathcal{X}} P_x^M \otimes V_x^X, \quad (15)$$

where $V_x^X |0\rangle^X = |x\rangle^X$. We can easily check that this unitary yields $\theta^{AX} = \text{Tr}_M[\theta^{AMX}]$, with θ^{AX} given by Eq. (6).

A. Measurement heat

Heat is one of the central quantities of thermodynamics [51], though its rigorous definition (along with the definition of internal energy) becomes problematic in certain regimes, such as in the presence of strong coupling [53,54] (see Appendix A). However, in the weak-coupling regime, where contributions to the internal energy due to the interaction Hamiltonian can be neglected, the following definition is accepted: for a system S of interest coupled with a thermal bath B , the heat $Q_t^{B \rightarrow S}$ absorbed by S from B or, equivalently, the heat $Q_t^{S \rightarrow B}$ dissipated from S into B is given by

$$-Q_t^{B \rightarrow S} \equiv Q_t^{S \rightarrow B} = \Delta \langle \mathcal{H}^B \rangle := \langle \mathcal{H}^B \rangle_t - \langle \mathcal{H}^B \rangle_0, \quad (16)$$

where $\langle \mathcal{H}^B \rangle_0 = \text{Tr}[\tau^B \mathcal{H}^B]$ and $\langle \mathcal{H}^B \rangle_t = \text{Tr}[\rho_t^B \mathcal{H}^B]$ denote the expectation of the bath Hamiltonian $\mathcal{H}^B$ at

the initial and final times, respectively. In Appendix A, we review the definitions of work and heat in more detail. There, we recall that under a set of three assumptions—(i) the interaction between S and B is unitary; (ii) B is initialized in a thermal state τ^B of inverse temperature $\beta = 1/k_B T$; and (iii) systems S and B are initially uncorrelated—one has [55]

$$\beta Q_t^{S \rightarrow B} = -\Delta H(S) + I(S : B) + D(\rho_t^B \| \tau^B). \quad (17)$$

There have been various efforts to characterize the heat absorbed or dissipated during a quantum measurement process (see, e.g., Refs. [5,6,23]). Here, we propose a definition of measurement heat of a quantum instrument $\{\mathcal{N}_x^A\}_{x \in \mathcal{X}}$ in the special case where the measurement system (probe) M is initially prepared in a thermal state, which is the natural choice when one is interested in the thermodynamics of quantum measurements (see, e.g., Refs. [28,29]). For that, recall that the quantum instrument formalism of a general quantum measurement can be described as a fully unitary interaction model between the system A being measured, the measurement apparatus M , and a classical memory X , as described in Eq. (14). Therefore, if the measurement apparatus M is initially in a thermal state of temperature T_m , then the assumptions leading to Eq. (17) apply, where the joint system AX plays the role of S , while M plays the role of the bath B . Accordingly, we have

$$\begin{aligned} \beta Q_{\text{meas}} &:= \beta Q^{M \rightarrow AX} \\ &= \Delta H(AX) - I(AX : M)_\theta - D(\theta^M \| \tau^M) \\ &= H(AX)_\theta - H(A)_\rho - I(AX : M)_\theta - D(\theta^M \| \tau^M) \\ &= H(X)_\theta - I_G - I(AX : M)_\theta - D(\theta^M \| \tau^M), \end{aligned} \quad (18)$$

where $I_G \equiv I_G(\{\mathcal{N}_x^A\}_{x \in \mathcal{X}}; \rho^A)$ is the Groenewold information gain computed for the premeasurement state ρ^A . From the above, we immediately see that a positive Groenewold information gain reduces the heat absorbed by AX from the measurement apparatus, whereas a negative Groenewold information gain increases that heat.

A tighter characterization can be given if $\theta^M = \tau^M$, that is, if the final reduced state of the measurement apparatus remains thermal. This is a plausible assumption for a macroscopic measurement apparatus, the internal state of which is not affected by the interaction with the system and the memory. In such a case, $D(\theta^M \| \tau^M) = 0$. We can then write

$$\beta Q_{\text{meas}} = \Delta H(AX) - I(AX : M)_\theta \quad (19)$$

$$=: \beta Q_{\text{meas}}^X + \beta Q_{\text{meas}}^{A|X}, \quad (20)$$

where we use the chain rule of quantum mutual information [39] and the definitions

$$\beta Q_{\text{meas}}^X := \Delta H(X) - I(X : M)_\theta = H(X|M)_\theta, \quad (21)$$

$$\beta Q_{\text{meas}}^{A|X} := \Delta H(A|X) - I(A : M|X)_\theta, \quad (22)$$

where $\Delta H(X) = H(X)_\theta$ because the classical memory X is initialized from the pure state $|0\rangle\langle 0|^X$. A less restrictive assumption is when the separation of $D(\theta^M \| \tau^M)$ into two parts is possible, corresponding to the two stages of evolution of M in Eq. (14), namely $\tau^M \rightarrow \omega^M$ and $\omega^M \rightarrow \theta^M$ where $\omega^M = \text{Tr}_A[U^{AM}(\rho^A \otimes \tau^M)(U^{AM})^\dagger]$. Then, under the assumption that $\text{Tr}[(\theta^M - \omega^M)(\ln \omega^M - \ln \tau^M)] = 0$, the Pythagorean theorem for relative entropy holds [56], i.e., $D(\theta^M \| \tau^M) = D(\theta^M \| \omega^M) + D(\omega^M \| \tau^M)$. The first and the second terms can be included in the definitions of $Q_{\text{meas}}^{A|X}$ and Q_{meas}^X , respectively. The above assumption, including τ^M , ω^M , and θ^M , need not be regarded as very restrictive for a macroscopic apparatus M , since the “errors” $(\theta^M - \omega^M)$ and $(\ln \omega^M - \ln \tau^M)$ get multiplied (for an exact statement, see Appendix B). Although we use $\theta^M = \tau^M$ in what follows, the alternative definitions of measurement heat contributions that include the relative entropy terms still preserve the results of this paper.

Note that, on the one hand, the quantity Q_{meas}^X is the amount of heat absorbed by the classical register X in a general quantum measurement (involving A , X , and M), which provides an alternative physical meaning to the conditional quantum entropy $H(X|M)_\theta$ described after Eq. (9), for the special case of a quantum measurement. On the other hand, since $H(A|X)_\rho = H(A)_\rho$, we can also write $Q_{\text{meas}}^{A|X}$ as

$$\beta Q_{\text{meas}}^{A|X} = -I_G - I(A : M|X)_\theta \leq -I_G, \quad (23)$$

where the latter does not depend on M . We summarize the above discussion as follows.

Theorem 1. *Suppose that a quantum system A undergoes a general quantum measurement described by the quantum instrument $\{\mathcal{N}_x^A\}_{x \in \mathcal{X}}$. Further, suppose that the measurement is physically implemented according to an indirect measurement $(\sigma^M, U^{AM}, \{P_x^M\}_{x \in \mathcal{X}})$, where σ^M (the internal state of the measurement apparatus) is thermal at temperature T_m . Suppose, moreover, that the reduced state of the measurement apparatus, after the measurement interaction, remains unchanged (it can, however, become correlated with the other systems). Then, the heat absorbed by the system A from the measurement apparatus M for each outcome $x \in \mathcal{X}$, averaged over all outcomes, is bounded from above by the negative of the Groenewold*

information gain:

$$Q_{\text{meas}}^{A|X} \leq -k_B T_m I_G. \quad (24)$$

Equivalently, the heat dissipated from the system into the measurement apparatus is bounded from below by the Groenewold information gain.

To our knowledge, this is the first attempt to provide a physical meaning to the Groenewold information gain even in the case when it is negative, which represents the vast majority (if not the totality) of all practically achievable quantum measurements. In fact, the Groenewold information gain is always non-negative only if the quantum measurement is *quasicomplete* [22], that is, when the postmeasurement states $\mathcal{N}_x^A(\rho^A)$ are all pure whenever ρ^A is pure: a condition that is arguably highly idealized.

IV. THERMODYNAMICS OF QEC ENGINES

In this section, we introduce the QEC engine and analyze its various stages of operation. The ultimate goal in this section is to derive a second law of thermodynamics for QEC engines that holds for an arbitrary initialization of the system, general (non-IID) thermal noise, and arbitrary quantum measurements, thereby generalizing results such as Ref. [27].

A. Engine components

We consider an arbitrary thermal operation $\mathcal{T}$ as the noisy channel that we would like to correct, namely

$$\mathcal{T}(\rho) = \text{Tr}_B [U(\rho \otimes \tau^B)U^\dagger], \quad (25)$$

where τ^B is a thermal state of the bath and U is a unitary operator that commutes with the total Hamiltonian [57]. We refer to the environment in this representation as the “hot” bath B_h with temperature T_h (inverse temperature β_h). We denote the system of interest and the purifying reference system by S and R , respectively. We also denote by A the ancillary system that is used to encode the state of the system.

In order to make a connection between QEC and heat engines with feedback, we need to introduce a cold bath B_c with temperature T_c (inverse temperature β_c). The purpose of the cold bath is to recycle the ancillary system after the QEC procedure, in preparation for future use. This raises the question of what the initial preparation of A should be for this construction to work. In the literature, the ancillary system is usually prepared in an initially pure state (often uncorrelated qubits) but this might not be possible for practical systems subject to thermal noise. Furthermore, resetting A back to a pure state [26] might imply that the cold bath has to operate at zero temperature $T_c = 0$,

which is not only unphysical but also detracts from the generality of possible thermodynamic statements that can be made for arbitrary T_c . Hence, to maintain full generality for arbitrary ancillary systems A and arbitrary temperatures T_c of the cold bath, we assume that A is prepared in an initially thermal state of the same temperature T_c . By denoting the energy basis of A by $\{|\varepsilon_x\rangle\}_x$, the initial state of the ancillary system is therefore given by

$$\tau_c^A = \sum_x p_X(x) |\varepsilon_x\rangle\langle\varepsilon_x|^A, \quad (26)$$

with

$$p_X(x) := e^{-\beta_c \varepsilon_x} / Z \quad \text{where} \quad Z = \sum_x e^{-\beta_c \varepsilon_x}. \quad (27)$$

In our paper, we accomplish the recycling back to the thermal state τ_c^A by conducting a SWAP operation between A and one of the uncorrelated subsystems of B_c that possesses the same dimensions and local Hamiltonian as A .

After initializing A in τ_c^A , we project it onto one of its energy eigenstates, in preparation for encoding. The outcome of the measurement is recorded using a classical register X , for the purpose of using it in the encoding and decoding stages of QEC. The last component in the QEC engine is the classical register Y that records the result of the error-syndrome measurement (conducted by a measurement apparatus M) and feeds it into the decoding channel, hence playing the same role as a feedback controller in a heat engine. For brevity, we use E for the combined engine components R, S, A, B_h , and B_c . To summarize, the QEC engine is comprised of the systems E, X , and Y , all of which are depicted in Fig. 2.

B. Cycle description

We now describe in detail the operation of the QEC engine presented in Fig. 2.

1. Initialization

The initial state of the engine is assumed to have a product form:

$$\sigma_i^{EXY} := \psi^{RS} \otimes \tau_c^A \otimes \tau_h^{B_h} \otimes \tau_c^{B_c} \otimes |0\rangle\langle 0|^X \otimes |0\rangle\langle 0|^Y, \quad (28)$$

where $\psi^{RS} = |\psi\rangle\langle\psi|^{RS}$ is a purification of the arbitrary initial system state σ_i^S , $|0\rangle\langle 0|^X$ and $|0\rangle\langle 0|^Y$ are the initial states of the classical registers X and Y , respectively, and τ_c^A , $\tau_h^{B_h}$, and $\tau_c^{B_c}$ are thermal states of A , B_h , and B_c with temperatures T_c , T_h , and T_c , respectively.

2. Initial-measurement stage

Here, we prepare A for the encoding stage by making a projective measurement $\{P_x^A\}_{x \in \mathcal{X}}$ onto its energy basis

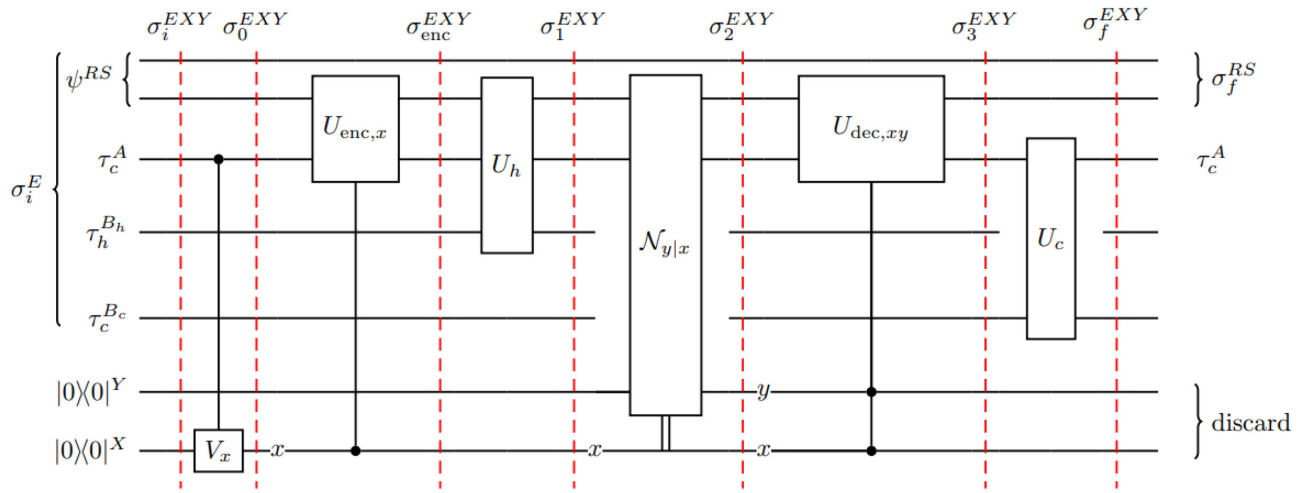


FIG. 2. The circuit description of the QEC engine. Lines locally disconnected from channels imply that the global channel (at that time step) has the effect of a local identity operation on the corresponding systems. We emphasize that the internal Hamiltonian of the reference system is assumed to be trivially zero at the initial and final times, namely $\mathcal{H}_i^R = \mathcal{H}_f^R \equiv 0$. All other (local and interaction) Hamiltonians are assumed to be cyclic, i.e., $\mathcal{H}_i = \mathcal{H}_f$. The only nonunitary steps in our engine are the error-syndrome measurement (characterized by the quantum instrument $\{\mathcal{N}_{y|x}^{SA}\}_y$ for a given x) and the discarding (memory erasure) of the classical registers X and Y .

(i.e., $P_x^A = |\varepsilon_x\rangle\langle\varepsilon_x|^A$) and hence the nonselective postmeasurement state is as follows:

$$\sigma_0^{EXY} := \sum_x p_X(x) \frac{P_x^A \sigma_i^E P_x^A}{p_X(x)} \otimes |x\rangle\langle x|^X \otimes |0\rangle\langle 0|^Y, \quad (29)$$

where

$$\frac{P_x^A \sigma_i^E P_x^A}{p_X(x)} = |\psi\rangle\langle\psi|^{RS} \otimes |\varepsilon_x\rangle\langle\varepsilon_x|^A \otimes \tau_h^{B_h} \otimes \tau_c^{B_c}. \quad (30)$$

3. Adaptive-encoding stage

Using the measurement outcome $x \in \mathcal{X}$, we encode the system RS using the ancillary system A and a unitary $U_{enc,x}^{SA}$ as follows:

$$\sigma_{enc,x}^{RSA} := U_{enc,x}^{SA} (|\psi\rangle\langle\psi|^{RS} \otimes |\varepsilon_x\rangle\langle\varepsilon_x|^A) (U_{enc,x}^{SA})^\dagger, \quad (31)$$

and the total state σ_{enc}^{EXY} becomes

$$\sum_x p_X(x) \sigma_{enc,x}^{RSA} \otimes \tau_h^{B_h} \otimes \tau_c^{B_c} \otimes |x\rangle\langle x|^X \otimes |0\rangle\langle 0|^Y. \quad (32)$$

4. Thermal-operation stage

The encoded system RSA interacts with the hot bath via the unitary U^{SAB_h} and leads to the following state:

$$\sigma_{1,x}^{RSAB_h} := U^{SAB_h} (\sigma_{enc,x}^{RSA} \otimes \tau_h^{B_h}) (U^{SAB_h})^\dagger. \quad (33)$$

Hence σ_1^{EXY} at this stage is given by

$$\sum_x p_X(x) \sigma_{1,x}^{RSAB_h} \otimes \tau_c^{B_c} \otimes |x\rangle\langle x|^X \otimes |0\rangle\langle 0|^Y. \quad (34)$$

It is important to note that, in general, the noisy channel does not act on S and A independently. This relaxes one of the assumptions often made in the QEC literature, namely that of independent noisy channels acting on individual subsystems (specifically qubits).

5. Error-syndrome-measurement stage

After the thermal-operation stage, we need a decoding process to recover the initial state of the system S . This process can be divided into two stages: error-syndrome measurement and error correction, characterized by the (x -value-dependent) quantum instrument $\{\mathcal{N}_{y|x}^{SA}\}_y$ and unitary operators $\{U_{dec,xy}^{SA}\}_{x,y}$, respectively, where $y \in \mathcal{Y}_x$ is the measurement outcome of the former stage (we write $\mathcal{Y}_x$ because the set of possible outcomes is generally a function of the initial-measurement outcome x). Therefore, the former is picked depending on the initial-measurement outcome $x \in \mathcal{X}$ and the latter can generally depend on both measurement outcomes $(x, y) \in (\mathcal{X}, \mathcal{Y}_x)$. The state of the engine after the error-syndrome-measurement step is described by the quantum-classical state [using the definition in Eq. (6)]

$$\sigma_2^{EXY} := \sum_{x,y} p_{XY}(x, y) \sigma_{2,xy}^E \otimes |x, y\rangle\langle x, y|^{XY}, \quad (35)$$

where $p_{XY}(x, y) = p_X(x) p_{Y|x}(y|x)$ and

$$\sigma_{2,xy}^E := \frac{\mathcal{N}_{y|x}^{SA}(\sigma_{1,x}^{RSAB_h})}{p_{Y|x}(y|x)} \otimes \tau_c^{B_c}. \quad (36)$$

It is important to note that the system Y is the classical register used to record the measurement outcome, not the measurement apparatus itself. The latter is implicit in the quantum instrument $\{\mathcal{N}_{y|x}^{SA}\}_y$ [25,37], which makes this stage nonunitary in general. Let us use the notation

$$\mathcal{N}_{|x}^{SA \rightarrow SAY}(\cdot) := \sum_{y \in \mathcal{Y}_x} \mathcal{N}_{y|x}^{SA}(\cdot) \otimes |y\rangle\langle y|^Y, \quad (37)$$

to indicate the joint action of the instrument on the system and the memory.

6. Error-correction stage

The decoding is completed by selectively applying $U_{\text{dec},xy}^{SA}$ based on the measurement outcomes (x,y) as

$$\sigma_3^{EXY} = \sum_{x,y} p_{XY}(x,y) \sigma_{3,xy}^E \otimes |x,y\rangle\langle x,y|^{XY}, \quad (38)$$

where

$$\sigma_{3,xy}^E = U_{\text{dec},xy}^{SA} \sigma_{2,xy}^E (U_{\text{dec},xy}^{SA})^\dagger. \quad (39)$$

7. Resetting of the ancillary system

To prepare the ancillary system for a future cycle, we let A interact with the cold bath to return it to its initial thermal state. In order to disentangle A from the rest of the engine, we pick the thermal operation at this stage to be a fully thermalizing channel, which simply replaces the final state of A with a newly prepared copy of τ_c^A . The unitary that realizes this process evolves the state via

$$\sigma_f^{EXY} = U^{AB_c} \sigma_3^{EXY} (U^{AB_c})^\dagger. \quad (40)$$

We note that this can be realized by taking $\sigma_3^{B_c} = \tau_c^{B_c} = \otimes_{i=1}^n \tau_c^{B_c^{(i)}}$ (for some integer n describing the size of the bath) and applying a SWAP operation (defined by $U_{\text{swap}}|\psi\rangle \otimes |\phi\rangle = |\phi\rangle \otimes |\psi\rangle$ for any two states $|\psi\rangle$ and $|\phi\rangle$) between σ_3^A and $\tau_c^{B_c^{(1)}}$, leading to $\sigma_f^A = \tau_c^A$ and $\sigma_f^{B_c} = \sigma_3^A \otimes_{i=2}^n \tau_c^{B_c^{(i)}}$.

8. Discarding the classical registers

At the final stage, we discard the classical registers X and Y . As we discuss after Eq. (9), this discarding process is equivalent to an erasure of the memory of Maxwell's demon, which is important for completing the thermodynamic cycle.

C. Information-theoretic analysis

In this subsection, we compute the entropy change at each stage of operation, with the purpose of arriving at

a formula for the net entropy change of the engine after discarding the classical registers X and Y . All information-theoretic quantities are computed for the sequence of states with subscripts

$$i \rightarrow 0 \rightarrow \text{enc} \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow f.$$

We start by noting that $H(EXY)_{\sigma_0} = H(EXY)_{\sigma_i}$ because σ_0 arises from σ_i by means of a controlled unitary, where A is the control system and X is the target system. The controlled unitary is given by its action

$$|\varepsilon_x\rangle^A \otimes |0\rangle^X \longrightarrow |\varepsilon_x\rangle^A \otimes |x\rangle^X, \quad (41)$$

via $V_x^X |0\rangle^X = |x\rangle^X$, and thus it can be written as

$$\sum_x |\varepsilon_x\rangle\langle \varepsilon_x|^A \otimes V_x^X. \quad (42)$$

The entropy at the encoding stage [defined by Eq. (32)] also does not change, i.e., $H(EXY)_{\sigma_{\text{enc}}} = H(EXY)_{\sigma_0}$, because this stage is also described by a controlled unitary, where X is the control system and RS_A is the target system. Moreover, it is clear from Eqs. (32) and (34) that $H(EXY)_{\sigma_1} = H(EXY)_{\sigma_{\text{enc}}}$ due to the unitary interaction between RS_A and B_h .

The first nonunitary step is the error-syndrome-measurement stage. The entropy after the measurement is computed as follows:

$$H(EXY)_{\sigma_2} = H(X) + H(EY|X)_{\sigma_2} \quad (43)$$

$$= H(X) + \sum_x p_X(x) H(EY)_{\sigma_{2,x}}, \quad (44)$$

where we use the notation $H(EY|X=x)_{\sigma_2} \equiv H(EY)_{\sigma_{2,x}}$. Note that it is convenient to isolate the contributions due to X because X is the system conditioned on which the measurement on the engine is chosen. We continue writing $H(EY)_{\sigma_{2,x}}$ in the form

$$\begin{aligned} H(EY)_{\sigma_{2,x}} &= H(E)_{\sigma_{1,x}} - [H(E)_{\sigma_{1,x}} - H(EY)_{\sigma_{2,x}}] \\ &= H(E)_{\sigma_{1,x}} - [H(E)_{\sigma_{1,x}} - H(Y)_{\sigma_{2,x}} - H(E|Y)_{\sigma_{2,x}}] \\ &= H(E)_{\sigma_{1,x}} + H(Y)_{\sigma_{2,x}} - I_G(\{\mathcal{N}_{y|x}^{SA}\}_y; \sigma_{1,x}^E), \end{aligned} \quad (45)$$

where in the last line we use the definition of the Groenewold information gain in Eq. (8). From now on, we denote the Groenewold information gain $I_G(\{\mathcal{N}_{y|x}^{SA}\}_y; \sigma_{1,x}^E)$ by $I_{G,x}$ for brevity. Substitution of this result back into

Eq. (44) yields

$$\begin{aligned} H(X) + \sum_x p_X(x) [H(E)_{\sigma_{1,x}} + H(Y)_{\sigma_{2,x}} - I_{G,x}] \\ = H(X) + H(E|X)_{\sigma_1} + H(Y|X)_{\sigma_2} - \sum_x p_X(x) I_{G,x} \\ = H(EXY)_{\sigma_1} + H(Y|X)_{\sigma_2} - \sum_x p_X(x) I_{G,x}, \end{aligned} \quad (46)$$

where in the last equality we use the fact that $H(EXY)_{\sigma_1} = H(EX)_{\sigma_1}$. Thus, the entropy change in the error-syndrome-measurement stage becomes

$$H(EXY)_{\sigma_2} - H(EXY)_{\sigma_1} = H(Y|X)_{\sigma_2} - \sum_x p_X(x) I_{G,x}. \quad (47)$$

Regarding the error-correction stage, we can easily see that $H(EXY)_{\sigma_3} = H(EXY)_{\sigma_2}$ because this stage is also described by a controlled unitary. Finally, we obviously have $H(EXY)_{\sigma_f} = H(EXY)_{\sigma_3}$ for the recycling stage of the ancillary system due to the unitarity of the SWAP operation between A and an identical subsystem from B_c .

The final step of the QEC cycle is the second and last nonunitary step after the error-syndrome measurement: here, we discard the classical registers X and Y . The entropy change due to the discarding process is simply given by $H(E)_{\sigma_f} - H(EXY)_{\sigma_f} = -H(XY|E)_{\sigma_f}$. Namely, discarding the classical registers only changes the entropy of the engine if the information encoded in them cannot be perfectly recovered from the final reduced state of the engine E .

From the above analysis, we can easily compute the total entropy change of E by using Eqs. (28) and (47) and we summarize it in the following.

Theorem 2. *The entropy change of the quantum error-correcting engine described in Fig. 2 after discarding the classical registers X and Y can be written in terms of information-theoretic quantities as*

$$\begin{aligned} H(E)_{\sigma_f} - H(E)_{\sigma_i} \\ = H(Y|X)_{\sigma_2} - \sum_x p_X(x) I_{G,x} - H(XY|E)_{\sigma_f}. \end{aligned} \quad (48)$$

D. Thermodynamic analysis

Here, we derive a formula for the entropy change $H(E)_{\sigma_f} - H(E)_{\sigma_i}$ in terms of the heat dissipated into the two baths (which can be measured as described in, e.g., Ref. [32]). If we denote by $\mathcal{H}^{B_{h(c)}}$ the Hamiltonian of the hot (cold) bath and $Q_{B_{h(c)}} = \langle \mathcal{H}^{B_{h(c)}} \rangle_i - \langle \mathcal{H}^{B_{h(c)}} \rangle_f$ the total heat absorbed by SA from the hot (cold) bath over the whole QEC cycle (where $\langle \cdots \rangle_i$ and $\langle \cdots \rangle_f$ denote

expectations over the initial and final states of the baths, respectively), then we can formulate the following theorem to describe the entropy change in terms of the dissipated heat.

Theorem 3. *The entropy change of the quantum error-correcting engine described in Fig. 2 after discarding the classical registers X and Y can be written in terms of the heat dissipated into the hot and cold baths as*

$$\begin{aligned} H(E)_{\sigma_f} - H(E)_{\sigma_i} = S_e - \frac{Q_h}{k_B T_h} - \frac{Q_c}{k_B T_c} \\ - I(RS : B_h B_c)_{\sigma_f} - \Gamma(B_h B_c)_{\sigma_f}, \end{aligned} \quad (49)$$

where

$$\begin{aligned} \Gamma(B_h B_c)_{\sigma_f} := I(B_h : B_c)_{\sigma_f} + D(\sigma_f^{B_h} \| \tau_h^{B_h}) \\ + D(\sigma_f^{B_c} \| \tau_c^{B_c}) \geq 0, \end{aligned} \quad (50)$$

depends only on the final states of the thermal baths and $S_e = H(RS)_{\sigma_f}$ is the entropy exchange.

Proof. We start from the final entropy of the engine

$$H(E)_{\sigma_f} \equiv H(RSAB_h B_c)_{\sigma_f} \quad (51)$$

$$= H(RSB_h B_c)_{\sigma_f} + H(A)_{\sigma_f}, \quad (52)$$

where we use the fact that the recycling of the ancillary system disentangles it from the rest of the engine, so that $I(RSAB_h B_c : A)_{\sigma_f} = 0$ [58]. Furthermore, using the definition of quantum mutual information, we can simplify $H(RSB_h B_c)_{\sigma_f}$ further, to

$$\begin{aligned} H(RSB_h B_c)_{\sigma_f} = H(RS)_{\sigma_f} + H(B_h)_{\sigma_f} + H(B_c)_{\sigma_f} \\ - I(B_h : B_c)_{\sigma_f} - I(RS : B_h B_c)_{\sigma_f}. \end{aligned} \quad (53)$$

Hence, for the entropy difference, we have

$$\begin{aligned} H(E)_{\sigma_f} - H(E)_{\sigma_i} = H(RS)_{\sigma_f} + \Delta H(B_h)_{\sigma_f} \\ + \Delta H(B_c)_{\sigma_f} - I(RS : B_h B_c)_{\sigma_f} \\ - I(B_h : B_c)_{\sigma_f}, \end{aligned} \quad (54)$$

where we use Eq. (28) and the fact that $\Delta H(RS) = H(RS)_{\sigma_f}$ because $\sigma_i^{RS} = |\psi\rangle\langle\psi|^{RS}$ is pure. Finally, we can use the Reeb-Wolf formula given in Eq. (17) for the

entropy changes of the two baths, yielding

$$\begin{aligned} H(E)_{\sigma_f} - H(E)_{\sigma_i} &= H(RS)_{\sigma_f} - \frac{Q_h}{k_B T_h} - \frac{Q_c}{k_B T_c} \\ &\quad - D(\sigma_f^{B_h} \| \tau_h^{B_h}) - D(\sigma_f^{B_c} \| \tau_c^{B_c}) \\ &\quad - I(RS : B_h B_c)_{\sigma_f} - I(B_h : B_c)_{\sigma_f}, \end{aligned} \quad (55)$$

which proves the theorem. $\blacksquare$

The entropy exchange S_e , introduced in Refs. [46,47], quantifies the purity of the final (error-corrected) system state σ_f^{RS} . Note that having $S_e = 0$ is only necessary, but not sufficient, is necessary for perfect QEC. More precisely, the entropy exchange is related to the entanglement fidelity given in Eq. (10) via the quantum Fano inequality, that is [47],

$$H(F_e) + (1 - F_e) \ln(d^2 - 1) \geq S_e, \quad (56)$$

where $H(x) := -x \log x - (1 - x) \log(1 - x)$ denotes the binary entropy and $d = \dim(\mathcal{H}_S \otimes \mathcal{H}_R)$. We can clearly see that $F_e = 1 \Rightarrow S_e = 0$ but the converse statement does not necessarily hold.

E. Second-law inequality

Using the two forms of the entropy change of E from Eqs. (48) and (49), we arrive at the total entropic balance equation for the QEC engine:

$$\begin{aligned} \frac{Q_h}{k_B T_h} + \frac{Q_c}{k_B T_c} &= S_e + \sum_x p_X(x) I_{G,x} \\ &\quad - H(Y|X)_{\sigma_2} + H(XY|E)_{\sigma_f} \\ &\quad - \Gamma(B_h B_c)_{\sigma_f} - I(RS : B_h B_c)_{\sigma_f}. \end{aligned} \quad (57)$$

The corresponding second-law inequality follows from $I(RS : B_h B_c)_{\sigma_f} \geq 0$ and $\Gamma(B_h B_c)_{\sigma_f} \geq 0$ and is stated as follows.

Corollary 1. *Given a general quantum error-correcting engine described in Fig. 2, the Clausius formulation of the second law takes the form*

$$\begin{aligned} \frac{Q_h}{k_B T_h} + \frac{Q_c}{k_B T_c} &\leq S_e + \sum_x p_X(x) I_{G,x} \\ &\quad - H(Y|X)_{\sigma_2} + H(XY|E)_{\sigma_f}, \end{aligned} \quad (58)$$

where the bound is achieved when $I(RS : B_h B_c)_{\sigma_f} = \Gamma(B_h B_c)_{\sigma_f} = 0$ (see the statement of Theorem 3).

Here, we remind the reader that a physical interpretation of each of the terms appearing in the second-law inequality

is provided above. This inequality adds upon the previous literature [26,27] in various ways. We elaborate further on this point in Sec. VI.

F. Applicability to multiple cycles

Finally, we comment in passing on extending this procedure to multiple rounds of QEC. This is natural to consider since, by taking into account the resetting of the classical memories and the ancillary system, we are preparing the QEC engine for a future cycle. A similar analysis can then be conducted for the follow-up cycle, provided that two conditions apply: (1) the remnant correlations between S and the heat baths are negligible; and (2) the change in the thermality of the baths can be ignored. Except for these considerations, we also need to recycle the reference system R in order to use it for purification in the new cycle. This costs no energy because we assume that R has a trivial Hamiltonian. Note that the measurement apparatus need not be reset under the assumption that its state remains approximately unchanged postmeasurement. Under these conditions, multiple-round QEC becomes possible. Some additional “resetting” time might be permitted in order for the baths B_h and B_c and the apparatus M to reach local equilibrium again.

V. EFFICIENCY-FIDELITY TRADE-OFF

In this section, we choose to focus on the thermodynamic efficiency of the QEC engine in Fig. 2. This is motivated by the following observation: a heat engine operating cyclically (with respect to both the state and the Hamiltonian of the working fluid) between two heat baths can possess a thermodynamic efficiency η that is no larger than the Carnot efficiency η_C (in the absence of a feedback controller). However, the state-cyclicity condition in heat-engine thermodynamics can be reinterpreted as a perfect QEC condition of the working fluid. In other words, maximum work extraction in thermodynamics can be viewed as the optimization of the efficiency $\eta \rightarrow \eta^* \equiv \max \eta$ of the engine (where $\eta^* \equiv \eta_C$ with no feedback), given the cyclicity constraint $F_e = 1$ in terms of the error-correction fidelity. Therefore, a natural extension to this observation is to ask whether the task of QEC can be framed in a similar way, namely as an optimization of the error-correction fidelity $F_e \rightarrow F_e^* \equiv \max F_e$, given any fixed value of the thermodynamic efficiency η within its allowable range. Additionally, recall that the QEC engine in Fig. 2 is a heat engine with a feedback controller and, therefore, the thermodynamic efficiency η can generally become larger than the Carnot efficiency η_C [27]. This means that the optimization of F_e^* for a thermodynamically desirable efficiency $\eta \geq \eta_C$ is of interest.

In this section, we derive a direct trade-off relation between F_e^* and η for a QEC engine operating in the super-Carnot regime ($\eta \geq \eta_C$). The (F_e^*, η) trade-off relation is a

special case of a triple-trade-off relation between the maximum achievable entanglement fidelity F_e^* , the thermodynamic efficiency η of the QEC engine, and the efficacy of the error-syndrome measurement. The triple-trade-off relation is derived under a set of physically motivated assumptions, the validity of which is further demonstrated in Sec. VF. As a consequence, a second trade-off relation exists between the efficiency and the efficacy [defined in Eq. (2)] whenever the state-cyclicity condition $F_e = 1$ (i.e., perfect QEC) is satisfied.

A. Origins of the trade-off: Quantum Fano inequality

In Sec. IVD, we discuss the quantum Fano inequality. This inequality can be rearranged in the form

$$H(F_e) \geq \ln(d^2 - 1)(F_e - 1) + S_e, \quad (59)$$

where the binary entropy on the left-hand side is a concave function in the region $F_e \in [0, 1]$, taking values between 0 (for $F_e \in \{0, 1\}$) and 1 (for $F_e = 0.5$) and the right-hand side is a linear function of F_e with a positive slope. A simple pictorial interpretation of the above inequality shows that for varying values of entropy exchange $S_e \geq 0$, there exists a direct trade-off relation between the exact value of S_e and the upper value F_e^* of the entanglement fidelity $F_e \in [0, F_e^*]$ satisfying the quantum Fano inequality for a fixed S_e . This observation constitutes the basis of the trade-off theorem that we derive in this section.

B. Definition of thermodynamic efficiency of a QEC engine

Here, we define what we mean by the thermodynamic efficiency of the QEC engine, noting that this requires a careful accounting of the input heat from the measurement apparatus, which we consider in Sec. III A. Such a problem can also be related to the problem of determining the thermodynamic efficiency of a heat engine operating between multiple heat baths (here, the measurement apparatus plays the role of the third bath).

We define the thermodynamic efficiency of our QEC engine in Fig. 2 as $\eta := -W_{\text{tot}}/Q_{\text{input}}$, where W_{tot} is the total work output of the engine and Q_{input} is the net positive heat that is absorbed by the “working fluid” SA . In what follows, we find an explicit form of Q_{input} . To do so, we start by noting that the net change in the internal energy of the working fluid SA is given by

$$\Delta \mathcal{U}^{SA} = \Delta \mathcal{U}^{RSA} = \Delta \mathcal{U}^{RSAY}, \quad (60)$$

where the first equality follows from the fact that $\sigma_f^R = \sigma_i^R$ (see the note after Fig. 2), whereas the second equality follows from the fact that the memory of Y is erased after the QEC operation. Note that although it is not obvious how to compute the work and heat contributions of $\Delta \mathcal{U}^{SA}$

of the working fluid SA alone [5,6,23,54,59,60], $\Delta \mathcal{U}^{RSAY}$ does not suffer from the same problem. In fact, based on the stages described in Sec. IV B, we have

$$\mathcal{U}_0^{RSAY} - \mathcal{U}_i^{RSAY} = 0, \quad (61)$$

$$\mathcal{U}_{\text{enc}}^{RSAY} - \mathcal{U}_0^{RSAY} = W_{\text{enc}}, \quad (62)$$

$$\mathcal{U}_1^{RSAY} - \mathcal{U}_{\text{enc}}^{RSAY} = Q_h, \quad (63)$$

where we use the definitions of heat and work resulting from Eq. (12). For the error-syndrome-measurement stage, we have

$$\mathcal{U}_2^{RSAY} - \mathcal{U}_1^{RSAY} = \text{Tr}[(\sigma_2^{RSAY} - \sigma_1^{RSAY})\mathcal{H}^{RSAY}], \quad (64)$$

where $\mathcal{H}^{RSAY} = \mathcal{H}_{t_1}^{RSAY} = \mathcal{H}_{t_2}^{RSAY}$ is the total Hamiltonian of $RSAY$ at the beginning and end of the error-syndrome-measurement stage (at times t_1 and t_2 , respectively). Recalling the definition of heat in Eq. (16), we arrive at $\mathcal{U}_2^{RSAY} - \mathcal{U}_1^{RSAY} = W_{\text{meas}} + Q_{\text{meas}}$, with $Q_{\text{meas}} = \langle \mathcal{H}^M \rangle_1 - \langle \mathcal{H}^M \rangle_2$, where $\mathcal{H}^M$ is the Hamiltonian of the measurement apparatus. Finally, for the last three stages of operation of the QEC engine, we have

$$\mathcal{U}_3^{RSAY} - \mathcal{U}_2^{RSAY} = W_{\text{dec}}, \quad (65)$$

$$\mathcal{U}_f^{RSAY} - \mathcal{U}_3^{RSAY} = Q_c, \quad (66)$$

$$\mathcal{U}_{\text{erase}}^{RSAY} - \mathcal{U}_f^{RSAY} = \langle \mathcal{H}^Y \rangle_i - \langle \mathcal{H}^Y \rangle_f \equiv Q_{\text{erase}}^Y, \quad (67)$$

where $\mathcal{H}^Y$ is the Hamiltonian of the classical register Y and $\langle \cdots \rangle_f$ denotes the expectation with respect to the final reduced state of Y , i.e., $\sigma_f^Y = \text{Tr}_{EX}[\sigma_f^{EXY}] = \sum_x p_X(x) \sigma_{f,x}^Y$, with $\sigma_{f,x}^Y = \sum_{y \in \mathcal{Y}_X} p_{Y|X}(y|x) |y\rangle\langle y|^Y$. Consequently, the above discussion yields

$$\Delta \mathcal{U}^{RSAY} = W_{\text{tot}} + Q_{\text{tot}}, \quad (68)$$

where

$$W_{\text{tot}} \equiv W_{\text{enc}} + W_{\text{meas}} + W_{\text{dec}}, \quad (69)$$

$$Q_{\text{tot}} \equiv Q_h + Q_{\text{meas}} + Q_c + Q_{\text{erase}}^Y \quad (70)$$

are the total work and heat separations of the internal energy of $RSAY$. Note that although the total dissipated heat during the erasure of both memory systems X and Y is given by $Q_{\text{erase}} = Q_{\text{erase}}^X + Q_{\text{erase}}^Y$, the term Q_{erase}^X is irrelevant for the above thermodynamic analysis of $\Delta \mathcal{U}^{RSAY}$.

To identify Q_{input} , we further decompose $Q_{\text{meas}} = Q_{\text{meas}}^Y + Q_{\text{meas}}^{RSAY} - k_B T_m D(\sigma_2^M \| \tau_m^M)$ based on our discussion in Sec. III A, where σ_2^M is the postmeasurement state of the measurement apparatus M , and τ_m^M is its initial

thermal state, with temperature T_m . By definition, we have

$$\frac{Q_{\text{meas}}^Y}{k_B T_m} = \Delta H(Y|X) - I(Y : M|X)_{\sigma_2} \quad (71)$$

$$= H(Y|XM)_{\sigma_2} \geq 0, \quad (72)$$

whereas $Q_{\text{meas}}^{RSA|Y} \leq 0$ is typically satisfied for any meaningful QEC engine. This is because a measurement process is supposed to transfer entropy (and hence information) from the measured system to the measurement apparatus. This, combined with $Q_h \geq 0$, $Q_c \leq 0$ (since we are interested in the heat-engine regime of operation) and $Q_{\text{erase}}^Y \leq 0$ (which follows from the fact that the memory Y is initialized from, and erased back to, its ground state) leads to the decomposition of Q_{tot} into positive $Q_h + Q_{\text{meas}}^Y$ and negative $Q_{\text{meas}}^{RSA|Y} + Q_c + Q_{\text{erase}}^Y$ parts. Therefore, by definition, we have $Q_{\text{input}} := Q_h + Q_{\text{meas}}^Y$.

Next, we make a physical simplification in the error-correction setting by introducing the following assumption.

Assumption 1. *We consider the regime of error correction with “sufficiently high fidelity,” such that $|\Delta \mathcal{U}^{RSAY}| \ll Q_{\text{input}}$.*

By “sufficiently high fidelity,” we mean that the QEC fidelity $F_e \geq 1 - \varepsilon$, with $\varepsilon \ll (Q_{\text{input}}/2\|\mathcal{H}^S\|_\infty)^2$ with $\|\cdot\|_\infty$ denoting the infinity norm of an operator. This is justified by noting that $\Delta \mathcal{U}^{RSAY} = \Delta \mathcal{U}^{RS}$ along with

$$|\Delta \mathcal{U}^{RS}| = |\text{Tr}[(\sigma_f^{RS} - \sigma_i^{RS})\mathcal{H}^{RS}]| \quad (73)$$

$$\leq \|\sigma_f^{RS} - \sigma_i^{RS}\|_1 \cdot \|\mathcal{H}^{RS}\|_\infty \quad (74)$$

$$\leq 2\|\mathcal{H}^S\|_\infty \sqrt{\varepsilon}, \quad (75)$$

where the first inequality is a special case of Hölder’s inequality and the second follows from the inequality $1/2\|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho, \sigma)}$, holding for every two density matrices.

Consequently, the thermodynamic efficiency of the QEC engine has the form $\eta := -W_{\text{tot}}/Q_{\text{input}} \approx Q_{\text{tot}}/Q_{\text{input}}$, which follows from dividing both sides of Eq. (68) by Q_{input} and then using Assumption 1. Also, $\eta \leq 1$ follows from the definition of Q_{input} as the positive part of Q_{tot} , appearing in Eq. (68).

C. Useful thermodynamic and information-theoretic statements

With the above definition of thermodynamic efficiency, let us now compute the important thermodynamic quantity $Q_h/k_B T_h + Q_c/k_B T_c$ appearing in Theorem 3. From

$\eta = Q_{\text{tot}}/Q_{\text{input}}$, it follows that

$$|Q_c| = (1 - \eta)(Q_h + Q_{\text{meas}}^Y) - |Q_{\text{meas}}^{RSA|Y}| - |Q_{\text{erase}}^Y|, \quad (76)$$

and therefore

$$\begin{aligned} \frac{Q_h}{k_B T_h} + \frac{Q_c}{k_B T_c} &= \frac{Q_h}{k_B T_h} \left(\frac{\eta - \eta_C}{1 - \eta_C} \right) \\ &\quad + \frac{|Q_{\text{meas}}^{RSA|Y}| + |Q_{\text{erase}}^Y| - (1 - \eta)Q_{\text{meas}}^Y}{k_B T_c}, \end{aligned} \quad (77)$$

where $\eta_C = 1 - T_c/T_h$ is the Carnot efficiency.

One more relation that we need is the upper bound on the Groenewold information gain derived in Ref. [24]. This is given (for a specific realization $x \in \mathcal{X}$ of the initial-measurement stage) by

$$I_{G,x} \leq H(Y|X = x)_{\sigma_2} - D(\sigma_{1,x}^E \| \mathcal{N}_{|x}^\dagger \circ \mathcal{N}_{|x}(\sigma_{1,x}^E)), \quad (78)$$

where we use the notation $\mathcal{N}_{|x} \equiv \mathcal{N}_{|x}^{SA \rightarrow SAY}$ from Eq. (37). The relative-entropy term is not necessarily positive because the input $\mathcal{N}_{|x}^\dagger \circ \mathcal{N}_{|x}(\sigma_{1,x}^E)$ is not a density matrix. This can be seen by normalizing the positive semidefinite operator $\mathcal{N}_{|x}^\dagger \circ \mathcal{N}_{|x}(\sigma_{1,x}^E)$ by dividing by its trace, leading to [33]

$$\begin{aligned} D(\sigma_{1,x}^E \| \mathcal{N}_{|x}^\dagger \circ \mathcal{N}_{|x}(\sigma_{1,x}^E)) &= D(\sigma_{1,x}^E \| \tilde{\sigma}_{1,x}^E) \\ &\quad - \ln \text{Tr}[\mathcal{N}_{|x}^\dagger \circ \mathcal{N}_{|x}(\sigma_{1,x}^E)], \end{aligned} \quad (79)$$

where the first term is now non-negative (because $\tilde{\sigma}_{1,x}^E$ is a density matrix) and the second term is the negative log efficacy $\mathcal{E}(\mathcal{N}_{|x}; \sigma_{1,x}^E)$ of the quantum instrument $\mathcal{N}_{|x}$ with respect to the density matrix $\sigma_{1,x}^E$, as defined in Eq. (2). Using these relations, we arrive at

$$\begin{aligned} H(Y|X)_{\sigma_2} - \sum_{x \in \mathcal{X}} p_X(x) I_{G,x} \\ \geq \sum_{x \in \mathcal{X}} p_X(x) [-\ln \mathcal{E}(\mathcal{N}_{|x}; \sigma_{1,x}^E)]. \end{aligned} \quad (80)$$

D. Lower bound on entropy exchange

Due to the importance of entropy exchange S_e in arriving at a trade-off relation (as emphasized in Sec. V A), we express this quantity as a function of the thermodynamic efficiency η , defined above. To start, we make the following remark that is justified in an experimental setting, since access to cooling resources is typically limited.

Remark 1. *We treat the cold bath B_c as a shared cooling resource among various subsystems in the QEC setting.*

That is, B_c is not only used to recycle the ancillary system A but it can also be used to erase the memory of the classical registers X and Y .

This constitutes the basis for the following assumption.

Assumption 2. *Preerasure, the entropy of the memory X is negligible with respect to the entropy of Y , namely $H(X|E)_{\sigma_f} \ll H(Y|EX)_{\sigma_f}$.*

Here, we offer the following physical justification. Due to Remark 1, we expect the shared cold bath (amongst systems A , X , and Y) to be sufficiently cold to accomplish the erasure task of the classical memories. Therefore, we expect the entropy of X (which depends on the statistics of the initial-measurement stage) to be negligible relative to that of Y , which is determined by the statistics of the error-syndrome measurement, the latter being rich in most realistic conditions.

From the above two assumptions, it follows that in Theorem 2, we have the approximation

$$H(XY|E)_{\sigma_f} = H(X|E)_{\sigma_f} + H(Y|EX)_{\sigma_f} \quad (81)$$

$$\approx H(Y|EX)_{\sigma_f}. \quad (82)$$

Moreover, we can further simplify $H(Y|EX)_{\sigma_f}$ as

$$H(Y|EX)_{\sigma_f} = H(Y|RSAB_h B_c X)_{\sigma_f} \quad (83)$$

$$= H(Y|RSB_c X)_{\sigma_f} \quad (84)$$

$$= H(Y|RSAX)_{\sigma_3} \quad (85)$$

$$= H(Y|X)_{\sigma_2} - I(Y : RSA|X)_{\sigma_3}. \quad (86)$$

Given that we can achieve QEC with sufficiently high fidelity (Assumption 1), $\sigma_{3,x}^{RSA}$ should be approximately back to the postencoded state $\sigma_{\text{enc},x}^{RSA}$. It follows that $I(Y : RSA|X)_{\sigma_3} \ll H(Y|X)_{\sigma_2}$ and hence $H(Y|EX)_{\sigma_f} \approx H(Y|X)_{\sigma_2}$. Finally, recall that the latter term also appears in Landauer's bound (see, e.g., Eq. (17), for a given realization x), which puts a lower bound on the erasure heat of the memory Y , as $|Q_{\text{erase}}^Y| \geq k_B T_c H(Y|X)_{\sigma_2}$. This holds because the erasure process of Y , for a given value x of the initial-measurement stage, can be carried out unitarily. Averaging Eq. (17) over the realizations $x \in \mathcal{X}$, we arrive at the desired inequality. Therefore, we make the following assumption.

Assumption 3. *The Landauer bound is reachable approximately in experiments, namely $|Q_{\text{erase}}^Y| \approx k_B T_c H(Y|X)_{\sigma_2}$.*

From this, it follows that $H(XY|E)_{\sigma_f} \approx H(Y|X)_{\sigma_2} \approx |Q_{\text{erase}}^Y|/k_B T_c$. Substituting this into Theorem 2, we obtain

$\Delta H(E)_{\sigma_f} \approx H(Y|X)_{\sigma_2} - \sum_{x \in \mathcal{X}} p_X(x) I_{G,x} - |Q_{\text{erase}}^Y|/k_B T_c$. This, along with Theorem 3 and Eq. (77), gives us

$$S_e \geq \frac{Q_h}{k_B T_h} \left(\frac{\eta - \eta_C}{1 - \eta_C} \right) + \frac{|Q_{\text{meas}}^{RSA|Y}| - (1 - \eta) Q_{\text{meas}}^Y}{k_B T_c} + \left[H(Y|X)_{\sigma_2} - \sum_{x \in \mathcal{X}} p_X(x) I_{G,x} \right]. \quad (87)$$

A simple manipulation via $(1 - \eta) Q_{\text{meas}}^Y = (1 - \eta_C) Q_{\text{meas}}^Y - (\eta - \eta_C) Q_{\text{meas}}^Y$ yields

$$S_e \geq \frac{Q_{\text{input}}}{k_B T_h} \left(\frac{\eta - \eta_C}{1 - \eta_C} \right) + \frac{|Q_{\text{meas}}^{RSA|Y}| - (1 - \eta_C) Q_{\text{meas}}^Y}{k_B T_c} + \left[H(Y|X)_{\sigma_2} - \sum_{x \in \mathcal{X}} p_X(x) I_{G,x} \right]. \quad (88)$$

The middle term on the right-hand side of the above equality becomes positive when $|Q_{\text{meas}}^{RSA|Y}| \geq \frac{T_c}{T_h} Q_{\text{meas}}^Y$, which implies that the measurement process leads to a “sufficient” dissipation of heat from the encoded system to the measurement apparatus. Therefore, the last assumption that we make is given as follows.

Assumption 4. *The thermodynamic constraint $|Q_{\text{meas}}^{RSA|Y}| \geq \frac{T_c}{T_h} Q_{\text{meas}}^Y$ applies to the measurement process. The equivalent information-theoretic constraint is given by $I(RSA : M|XY)_{\sigma_2} + \sum_x p_X(x) I_{G,x} \geq \frac{T_c}{T_h} H(Y|X)_{\sigma_2}$.*

This assumption, along with Eq. (80), yields the final inequality:

$$S_e \geq \frac{Q_{\text{input}}}{k_B T_h} \left(\frac{\eta - \eta_C}{1 - \eta_C} \right) + \sum_{x \in \mathcal{X}} p_X(x) \left[-\ln \mathcal{E}(\mathcal{N}_{|x}; \sigma_{1,x}^E) \right]. \quad (89)$$

Therefore, we summarize the main result of this section as follows.

Theorem 4. *Given a quantum error-correcting engine shown in Fig. 2 and the set of physical Assumptions 1–4 provided above, the entropy exchange is bounded from below by the thermodynamic efficiency of the engine and the average negative log efficacy of the general quantum measurement implemented, namely*

$$S_e \geq \frac{Q_{\text{input}}}{k_B T_h} \left(\frac{\eta - \eta_C}{1 - \eta_C} \right) + \sum_{x \in \mathcal{X}} p_X(x) \left[-\ln \mathcal{E}(\mathcal{N}_{|x}; \sigma_{1,x}^E) \right]. \quad (90)$$

E. Derivation of a triple-trade-off relation

Finally, trade-off relations between entanglement fidelity, thermodynamic efficiency, and efficacy now appear as a direct consequence of the quantum Fano inequality discussion in Sec. V A and Theorem 4. We formulate this as the following triple-trade-off relation.

Corollary 2. *Given a quantum error-correcting engine shown in Fig. 2 and the set of physical Assumptions 1–4 provided above, there exists a triple trade-off between the maximum achievable error-correcting fidelity F_e^* , the super-Carnot efficiency η of this error-correcting engine, and the average negative log efficacy of the general quantum measurement, namely*

$$H(F_e) + (1 - F_e) \ln(d^2 - 1) \geq \frac{Q_{\text{input}}}{k_B T_h} \left(\frac{\eta - \eta_C}{1 - \eta_C} \right) + \sum_{x \in \mathcal{X}} p_X(x) [-\ln \mathcal{E}(\mathcal{N}_{|x}; \sigma_{1,x}^E)]. \quad (91)$$

From here, two corollaries follow.

Corollary 3. *Given a quantum error-correcting engine shown in Fig. 2 and the set of physical Assumptions 1–4 provided above, along with the assumption that the quantum instrument used has efficacy larger than one, there exists a direct trade-off between the maximum achievable error-correcting fidelity F_e^* and the super-Carnot efficiency η of this error-correcting engine, namely*

$$\eta > \eta_C \Rightarrow F_e^* < 1 \text{ or, equivalently, } F_e^* = 1 \Rightarrow \eta \leq \eta_C. \quad (92)$$

Corollary 4. *Given a quantum error-correcting engine shown in Fig. 2 and the set of physical Assumptions 1–4 provided above, along with the assumption that the entanglement fidelity of the engine is equal to one, there exists a direct trade-off between the super-Carnot efficiency η of the error-correcting engine, and the average negative log efficacy of the general quantum measurement, namely*

$$\frac{Q_{\text{input}}}{k_B T_h} \left(\frac{\eta - \eta_C}{1 - \eta_C} \right) + \sum_{x \in \mathcal{X}} p_X(x) [-\ln \mathcal{E}(\mathcal{N}_{|x}; \sigma_{1,x}^E)] \leq 0. \quad (93)$$

The latter implies that for all subunitary quantum instruments $\mathcal{N}_{|x}$ (which includes efficient measurements), we must have $\eta \leq \eta_C$. This is in contrast with Ref. [27]. In fact, operation in the super-Carnot efficiency regime is still physically possible; however, it necessarily requires the use of superunitary quantum instruments. Further comparison with Ref. [27] is made in Sec. VI.

F. Examples

1. Bit-flip channel

Here, we consider the simplest QEC code: the three-qubit code corresponding to the bit-flip noise channel $\mathcal{T}(\cdot) = (1 - p)(\cdot) + pX(\cdot)X$, where $0 \leq p \leq 1$ is the bit-flip probability [11]. This can model the so-called flip-flop interactions between spins. First, note that the bit-flip channel can be written as a thermal operation [see Eq. (25)]. This can be accomplished by having the energy spectrum of the hot bath to be comprised of two energies ϵ_0 and ϵ_1 (where the excited state can be highly degenerate), with projectors P_0 and P_1 , respectively. Hence we choose the thermal state of the bath to be $\tau_h^{B_h} = (1 - p)P_0 + pP_1$ and the unitary $U = I \otimes P_0 + X \otimes P_1$.

The system of interest S is a single qubit prepared in an arbitrary pure state $|\psi\rangle^S = a|0\rangle^S + b|1\rangle^S$ (hence there is no need for a purifying system R) and the ancillary system is two noninteracting qubits, each prepared in the thermal state $\tau_c^{A_1} = \pi_1|0\rangle\langle 0|^{A_1} + (1 - \pi_1)|1\rangle\langle 1|^{A_1}$ and $\tau_c^{A_2} = \pi_2|0\rangle\langle 0|^{A_2} + (1 - \pi_2)|1\rangle\langle 1|^{A_2}$, with Gibbs distributions $\{\pi_1, (1 - \pi_1)\}$ and $\{\pi_2, (1 - \pi_2)\}$ respectively, having the same temperatures T_c but slightly different energy separations (to avoid degeneracies in the combined ancillary system $A_1 A_2$). The initial projective measurement of the ancillary system A can have four different outcomes $\{00, 01, 10, 11\}$ with probabilities $\{\pi_1\pi_2, \pi_1(1 - \pi_2), (1 - \pi_1)\pi_2, (1 - \pi_1)(1 - \pi_2)\}$, respectively. Therefore, the encoding stage starts by applying the unitaries $\{I_1 \otimes I_2, I_1 \otimes X_2, X_1 \otimes I_2, X_1 \otimes X_2\}$ based on the corresponding measurement outcomes, so that we can arrive at the same postmeasurement state $|00\rangle\langle 00|^A$. After this step, we complete the unitary encoding by applying two controlled-NOT (CNOT) gates (one on each ancillary qubit), leading to

$$(a|0\rangle^S + b|1\rangle^S) \otimes |0\rangle^{A_1} \otimes |0\rangle^{A_2} \rightarrow a|000\rangle^{SA} + b|111\rangle^{SA} \equiv |\Psi\rangle^{SA}, \quad (94)$$

which is independent of x . Next, we subject the encoded system to the bit-flip channel, acting independently on each qubit, namely $\mathcal{T}^{\otimes 3}(|\Psi\rangle\langle\Psi|^{SA})$. We follow this step by conducting the error-syndrome measurement, described by the four projectors

$$\begin{cases} \Pi_{y=0}^{SA} = |000\rangle\langle 000|^{SA} + |111\rangle\langle 111|^{SA}, \\ \text{decoded by } U_{\text{dec}, y=0}^{SA} = I^{SA}, \\ \Pi_{y=1}^{SA} = |100\rangle\langle 100|^{SA} + |011\rangle\langle 011|^{SA}, \\ \text{decoded by } U_{\text{dec}, y=1}^{SA} = X^S \otimes I^{A_1} \otimes I^{A_2}, \\ \Pi_{y=2}^{SA} = |010\rangle\langle 010|^{SA} + |101\rangle\langle 101|^{SA}, \\ \text{decoded by } U_{\text{dec}, y=2}^{SA} = I^S \otimes X^{A_1} \otimes I^{A_2}, \end{cases}$$

$$\begin{cases} \Pi_{y=3}^{SA} = |001\rangle\langle 001|^{SA} + |110\rangle\langle 110|^{SA}, \\ \text{decoded by } U_{\text{dec}, y=3}^{SA} = I^S \otimes I^{A_1} \otimes X^{A_2}, \end{cases}$$

corresponding to no bit-flip error and a single bit-flip error in each qubit. Hence the measurement statistics are given by $p_{Y|X}(y|x) = p_Y(y) = \text{Tr} \left[\mathcal{T}^{\otimes 3}(|\Psi\rangle\langle\Psi|^{SA}) \Pi_y^{SA} \right]$ for $y = \{0, 1, 2, 3\}$. Therefore, we require a memory system Y with at least four orthonormal basis vectors $\{|y\rangle\}_{y=0}^3$ (e.g., a system of two qubits).

In order to check the validity of the assumptions used to derive Theorem 2, we need only check Assumptions 2 and 4, since Assumption 1 depends on the strength of the noise present, Assumption 3 has been validated experimentally [61], and the efficacy of the quantum instrument is equal to zero because the error-syndrome measurement is projective (and hence $\mathcal{N}_{|x}$ is unital). To check Assumption 2, we note that

$$H(X|E)_{\sigma_f} = H(X)_{\sigma_f} - I(X : E)_{\sigma_f} \quad (95)$$

$$= H(X)_{\sigma_f}, \quad (96)$$

$$H(Y|EX)_{\sigma_f} = H(Y|E)_{\sigma_f}. \quad (97)$$

Furthermore, we have

$$H(X)_{\sigma_f} = H(A)_{\tau_c}, \quad (98)$$

$$H(Y|E)_{\sigma_f} = H(Y)_{\sigma_f} - I(Y : SAB_h B_c)_{\sigma_f} \quad (99)$$

$$= H(Y)_{\sigma_2} - I(Y : SB_c)_{\sigma_f} \quad (100)$$

$$= H(Y)_{\sigma_2} - I(Y : SA)_{\sigma_3}, \quad (101)$$

which gives us an easy way to check Assumption 2, as $H(A)_{\tau_c} \ll H(Y)_{\sigma_2} - I(Y : SA)_{\sigma_3}$. For experimental convenience (controllability, ease of initialization, etc.), we assume that the qubit energies used are in the optical range, i.e., $\hbar\omega \sim 2 - 3$ eV, and we assume the worst-case scenario that the cold bath is at room temperature; hence $k_B T_c \approx 1/40$ eV. Consequently, we have for $\hbar\omega/k_B T_c \sim 100$. This makes $H(A)_{\tau_c} \sim 10^{-42}$ (which is a function of the ratio $\hbar\omega/k_B T_c$), which we verify to be much smaller than $H(Y)_{\sigma_2} - I(Y : SA)_{\sigma_3} = H(SAY)_{\sigma_3} - H(SA)_{\sigma_3} \approx 0.166$ for a bit-flip error probability of $p = 0.01$, when $a = b = 1/\sqrt{2}$. Furthermore, Assumption 4 is easily verified by checking that $I_G = H(Y|X)_{\sigma_2} = H(Y)_{\sigma_2} \approx 0.166$; hence $|Q_{\text{meas}}^{RSA|Y}| \geq \frac{T_c}{T_h} Q_{\text{meas}}^Y$ is true for any reasonable ratio T_c/T_h of the QEC engine.

2. Phase-damping channel

We would like to supplement the above example, which has a classical analog, with the phase-damping noise,

described by a single parameter λ as

$$\sigma = \begin{pmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{pmatrix} \rightarrow \begin{pmatrix} |a|^2 & ab^*e^{-\lambda} \\ a^*be^{-\lambda} & |b|^2 \end{pmatrix}. \quad (102)$$

Phase damping is a purely quantum-mechanical noise that appears in various implementations of qubits, e.g., nitrogen-vacancy centers, where there is a large mismatch between transition energies of the qubit and that of the bath spins [62,63] (hence flip-flop interactions are suppressed). One possible way to arrive at this noise via a thermal operation [see Eq. (25)] is to consider two systems, where the action of one on the other can be ignored (e.g., when one of the systems is macroscopic). Therefore, the larger system will be in local equilibrium, i.e., in the thermal state τ . Furthermore, the joint Hamiltonian of the two systems can be written as $\mathcal{H}_{\text{eff}}(t) + \mathcal{H}^B$, where $\mathcal{H}_{\text{eff}}$ is an effective stochastic Hamiltonian of the system. Averaging over all realizations of the stochastic noise, one can arrive at Eq. (102), where λ is generally a function of time.

It is known in the literature (see, e.g., Ref. [11]) that a phase-damping channel is equivalent to a phase-flip channel $\mathcal{T}(\cdot) = (1-p)(\cdot) + pZ(\cdot)Z$, where $0 \leq p \leq 1$ is the phase-flip probability. This equivalence follows from the unitary freedom in the choice of the Kraus operators of the phase-damping channel, as

$$\begin{aligned} & \left\{ \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\lambda} \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\lambda} \end{pmatrix} \right\} \\ & \rightarrow \left\{ \sqrt{\mu} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \sqrt{1-\mu} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right\}, \end{aligned} \quad (103)$$

where $\mu = (1 - \sqrt{1 - \lambda})/2$. Consequently, we need only correct the phase-flip channel. This can be done by encoding S using two ancillary qubits $A = A_1 A_2$ in the exact same way as for a bit-flip noise, with the difference that we include an additional step of applying a Hadamard gate to each of the three qubits [11], leading to the encoding

$$\begin{aligned} & (a|0\rangle^S + b|1\rangle^S) \otimes |0\rangle^{A_1} \otimes |0\rangle^{A_2} \\ & \rightarrow a|++\rangle^{SA} + b|--\rangle^{SA} \equiv |\Phi\rangle^{SA}, \end{aligned} \quad (104)$$

where $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. Therefore, we can continue in the exact same way as for the bit-flip case, but in the $|\pm\rangle$ basis, to arrive at the exact same results.

VI. COMPARISON WITH PREVIOUS RESULTS

Our main results can be summarized by Eqs. (24), (58), and (91). At this stage, we would like to show that Eq. (58) generalizes a result of Sagawa and Ueda [27]. First, note that the authors do not make the distinction between the main system S and the ancilla system A in their letter. As a

consequence, there is no implementation of an initial projective measurement to prepare the ancillary system and hence the classical register X is not present. More importantly, the authors do not discard the classical registers, which is equivalent to refraining from the erasure of the memory of Maxwell's demon (a practice that has historically led to many contradictions in thermodynamics, as we discuss in Sec. I). This leads to the absence of the important $H(XY|E)_{\sigma_f}$ term in the entropic balance leading to Eq. (58). Making these changes in the derivation, we arrive at

$$\frac{Q_h}{k_B T_h} + \frac{Q_c}{k_B T_c} \leq S_e + I_G, \quad (105)$$

which is the main result of Ref. [27], but with two important generalizations. (i) The initial state of the system is arbitrary in our setup (rather than thermal). (ii) The quantum measurement stage is completely general and the information gain from the measurement is quantified by the Groenewold information gain: this is in contrast with the special case of efficient measurements made in Ref. [27], where the information gain is quantified by the “quantum-classical mutual information,” which is a special case of the Groenewold information gain for efficient measurements.

A important attempt to derive the second law of QTD in the context of QEC has been made in Ref. [26]. There, the authors arrive at a formula for the total entropy production of a QEC engine. However, we consider this derivation to be too restrictive, for the following reasons. First, the authors consider a system of qubits acted upon independently by generalized amplitude damping noise (which is a special type of thermal noise). Second, the dissipated heat, work, and entropy production during the error-syndrome-measurement stage is neglected, whereas we show that it deserves careful consideration. Third, the ancillary system (also a collection of qubits) is initialized and recycled back to its ground state, implying the presence of a cold bath of temperature $T_c = 0$. The latter implication is not only unphysical but it also does not allow for the derivation of other forms of the second law that include arbitrary values of the cold-bath temperature, e.g., the Clausius formulation as Eq. (58) in our paper. Incidentally, it is exactly this issue that is circumvented by initializing the ancillary system from a thermal state rather than its ground state, including the classical register X , and conducting the initial-measurement stage before proceeding with the encoding, as seen in Fig. 2. Finally, it is worth mentioning that, similar to Ref. [27], the authors do not consider the entropic cost of erasing the memory of the classical registers in their final entropic balance.

An important approach for including the work costs of measurement and erasure processes in a heat engine with feedback control has been discussed in Ref. [64]. Indeed,

this prior contribution can be seen as complementary to our approach in Sec. III when the measurement on the system is conducted adiabatically, i.e., when no measurement heat is dissipated. A similar discussion can be found in Ref. [65] in the context of entanglement-driven quantum engines.

Finally, we would like to note that further insights into integrating the Groenewold information gain (when it is positive) and thermodynamics under a more unifying context of work extraction can be found in past literature (see, e.g., Refs. [23,65,66]).

VII. CONCLUSIONS AND OPEN QUESTIONS

In this paper, we take a thermodynamic point of view with respect to general quantum measurements (described in the quantum instrument formalism) and quantum error correction. The former allows us to arrive at an important bound on the measurement heat in terms of the Groenewold information gain. This bound grants a physical interpretation of negative values of the Groenewold information gain as the directionality of the dissipated heat that is physically allowed in a measurement process.

Moreover, we consider the thermodynamic approach to quantum error correction, regarding it as a heat engine with a feedback controller. This approach allows us to derive the second-law inequality for a QEC engine under very general conditions. As a manifestation of the second law in this setting, we show that a trade-off relation exists between the maximum achievable fidelity of the error-correction process and the attainability of super-Carnot efficiencies of a QEC engine.

An open question here is that of time, namely: how long does the QEC process take to be completed? Is there any similar trade-off relation between the time that it takes and the entanglement fidelity? This is directly related to the distinction between operating at maximum efficiency and maximum power for a heat engine. Extensions to multiple rounds of QEC will also be of practical interest in future works. Another interesting question is the manifestation of the second law in the strong coupling regime between the system and the hot bath. It is well known that, in this regime, there is no agreed-upon definition of heat and internal energy [53,54] and so a consideration of various definitions and their outcomes can be insightful.

We would finally like to mention that our efficiency-fidelity trade-off shares some similarities with thermodynamic uncertainty relations (TURs) [67–70]. This can serve as a second point of contact between thermodynamics and error correction.

ACKNOWLEDGMENTS

We would like to acknowledge Jonathan P. Dowling for his last course on quantum information and computing, which sparked the interest in this topic. A.D. would like to thank Armen Allahverdyan and Giacomo Guarnieri for

very useful discussions. This work was supported by the U.S. Army Research Office through the U.S. MURI Grant No. W911NF-18-1-0218. F.B. acknowledges support from the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT) Quantum Leap Flagship Program (MEXT Q-LEAP) Grant No. JPMXS0120319794 and the Japan Society for the Promotion of Science (JSPS) KAKENHI Grants No. 19H04066, No. 20K03746, and No. 21H05183. The LaTeX package QUANTIKZ [71] has been used to generate the circuit plot.

APPENDIX A: WORK AND HEAT

In this appendix, we define the thermodynamic quantities of work, internal energy, and heat in a general setting (both weak and strong coupling limits).

Consider two systems, A and B , and suppose that the collective system AB evolves unitarily in the time interval $[0, t]$ via the joint unitary V_t^{AB} . From the conservation of the total von Neumann entropy under unitary evolution $H(A, B)_{\sigma_t} = H(A, B)_{\sigma_0}$, we find the following for the sum of the local entropy changes:

$$\Delta H(A) + \Delta H(B) = \Delta I(A : B) \geq 0, \quad (\text{A1})$$

where we use the definition of conditional entropy. We postpone the standard assumption of $I(A : B)_{\sigma_0} = 0$ to arrive at more general results [52, 72, 73], since in practice one might not be able to initialize AB in a product form. Next, we relate the entropy change of B with $D(\sigma_t^B \| \sigma_0^B)$. This can be achieved as follows:

$$\Delta H(B) = -\text{Tr}_B [\sigma_t^B \ln \sigma_t^B] + \text{Tr}_B [\sigma_0^B \ln \sigma_0^B] \quad (\text{A2})$$

$$= -\text{Tr}_B [\sigma_t^B \ln \sigma_t^B] + \text{Tr}_B [\sigma_t^B \ln \sigma_0^B] - \text{Tr}_B [\sigma_t^B \ln \sigma_0^B] + \text{Tr}_B [\sigma_0^B \ln \sigma_0^B] \quad (\text{A3})$$

$$= \text{Tr}_B [(\sigma_0^B - \sigma_t^B) \ln \sigma_0^B] - D(\sigma_t^B \| \sigma_0^B). \quad (\text{A4})$$

The operator $\ln \sigma_0^B$ can be interpreted as an effective unitless Hamiltonian of the system B (i.e., $\beta \mathcal{H}_{\text{eff}}^B$ where β plays the role of inverse temperature); as a consequence, $\beta^{-1} D(\sigma_t^B \| \sigma_0^B)$ can be interpreted as the change in the free energy from the Gibbs state of B with the effective Hamiltonian. Combining Eqs. (A1) and (A4), we arrive at

$$\begin{aligned} \text{Tr}_B [(\sigma_0^B - \sigma_t^B) \ln \sigma_0^B] &= -\Delta H(A) + \Delta I(A : B) \\ &\quad + D(\sigma_t^B \| \sigma_0^B). \end{aligned} \quad (\text{A5})$$

Now, we make an assumption about controllability, thereby operationally distinguishing system A from B . Namely, we assume that system A can be controlled by an external parameter with predetermined dynamics $\{\lambda_t\}_{t \geq 0}$; this is in contrast with system B , which is assumed to be uncontrollable [51]. The assumption of controllability is

further extended to the interaction between A and B , since one might want to describe a measurement process or a thermalization process that has a coupling phase (with a measurement apparatus or a heat bath, respectively) and a decoupling phase. This can be written using the interaction Hamiltonian $\mathcal{V}^{AB}$, which depends on a second external control parameter $\{\nu_t\}_{t \geq 0}$ with predetermined dynamics. The controllability assumptions are reflected in the total (time-dependent) Hamiltonian as

$$\mathcal{H}_t^{AB} = \mathcal{H}_{\lambda_t}^A + \mathcal{H}^B + \mathcal{V}_{\nu_t}^{AB}, \quad (\text{A6})$$

with the interaction unitary given by

$$V_t^{AB} = \mathcal{T} \exp \left\{ -\frac{i}{\hbar} \int_0^t d\tau \mathcal{H}_\tau^{AB} \right\}, \quad (\text{A7})$$

where $\mathcal{T}$ denotes time ordering. Due to the fact that the collective system AB is isolated, the change in the total energy of the system is (by definition) equated to the work done on AB through the control parameters λ_t and ν_t as

$$W_t^{AB} = \int_0^t d\tau \text{Tr} [\sigma_\tau^{AB} (\partial_\tau \mathcal{H}_\tau^A + \partial_\tau \mathcal{V}_\tau^{AB})] \quad (\text{A8})$$

$$= \text{Tr} [\sigma_t^{AB} \mathcal{H}_t^{AB}] - \text{Tr} [\sigma_0^{AB} \mathcal{H}_0^{AB}]. \quad (\text{A9})$$

In the weak-coupling limit, this definition reduces to the local definition $W_t^A = \text{Tr} [\sigma_t^A \mathcal{H}_{\lambda_t}^A] - \text{Tr} [\sigma_0^A \mathcal{H}_{\lambda_0}^A]$ [51], which we use in this paper.

In order to define what we mean by heat, we need to specify what we mean by the internal energy $\mathcal{U}_t^A$ of the system A , at time t . To this end, there have been various approaches in the literature. The first is that internal energy has to be a local quantity; i.e., it has to be represented in terms of the reduced density matrix of the system under consideration [54]. The second approach is that the definition of internal energy need not be a local quantity and the heat dissipated from A should (by definition) be equal to the heat absorbed by B [53], i.e., $Q_t^A = -Q_t^B$ at all times. Both of these approaches can be used to derive the laws of thermodynamics and they become identical in the weak-coupling limit. In this paper, we adopt the second approach by using the same definition as in Ref. [53], in the form

$$\mathcal{U}_t^A = \text{Tr} [\sigma_t^{AB} (\mathcal{H}_{\lambda_t}^A + \mathcal{H}_{\nu_t}^{AB})], \quad (\text{A10})$$

which is clearly nonlocal in A and reduces to the well-known (local) definition $\text{Tr}_A [\sigma_t^A \mathcal{H}_{\lambda_t}^A]$ in the weak-coupling limit. The definition in Eq. (A10) is motivated by the intuition that the concept of an *internal* energy local to system A is unclear if its interaction with a second system B is not

negligible. From Eq. (A10), we easily obtain the first law of thermodynamics:

$$\Delta \mathcal{U}^A = W_t^{AB} + \langle \mathcal{H}^B \rangle_0 - \langle \mathcal{H}^B \rangle_t, \quad (\text{A11})$$

where $Q_t^A = -Q_t^B := \langle \mathcal{H}^B \rangle_0 - \langle \mathcal{H}^B \rangle_t$ is interpreted as the heat absorbed by A whenever B is a heat bath.

We consider the important case when system B is a heat bath. This is the most standard case in the literature, where we assume that system B is initially in a thermal state $\sigma_0^B = \tau^B = \exp(-\beta \mathcal{H}^B)/Z^B$. Then, Eq. (A5) gives

$$\beta \Delta \langle \mathcal{H}^B \rangle = -\Delta H(A) + \Delta I(A : B) + D(\sigma_t^B \| \tau^B), \quad (\text{A12})$$

which, combined with the assumption of an initial product state $\sigma_0^{AB} = \sigma_0^A \otimes \tau^B$, leads to the well-known Landauer bound $\beta Q_t^B \geq -\Delta H(A)$ [32,55,74]. System A in this context is usually identified with the memory system of Maxwell's demon (not to be confused with Maxwell's demon itself, which typically represents the measurement apparatus). Using Eqs. (A12) and (A11), we arrive at

$$\begin{aligned} \Delta \mathcal{U}^A &= W_t^{AB} + k_B T \Delta H(A) \\ &\quad - k_B T [\Delta I(A : B) + D(\sigma_t^B \| \tau^B)]. \end{aligned} \quad (\text{A13})$$

When the initial state of AB is of product form, we immediately obtain the familiar first law inequality of nonequilibrium thermodynamics [51]:

$$\Delta \mathcal{U}^A \leq W_t^{AB} + k_B T \Delta H(A), \quad (\text{A14})$$

where the equality is reached for a thermodynamically “reversible” process described by the conditions $I(A : B)_{\sigma_t} = 0$ and $D(\sigma_t^B \| \tau^B) = 0$. Furthermore, Thompson's formulation of the second law [75] follows almost immediately from the definition of free energy of A , as [53]

$$F_t^A := \mathcal{U}_t^A - k_B T H(A)_{\sigma_t}, \quad (\text{A15})$$

which, combined with Eq. (A14), yields

$$W_t^{AB} \geq \Delta F^A. \quad (\text{A16})$$

These rederivations of well-known thermodynamical relations further justify the definitions of internal energy $\mathcal{U}^A$ and thermodynamic work W^{AB} for arbitrary coupling strengths between systems A and B as used in Ref. [53].

APPENDIX B: APPROXIMATE PYTHAGOREAN THEOREM OF RELATIVE ENTROPY

Consider three density matrices, τ , ω , and θ , that are selected from $\mathcal{D}(\mathcal{H})$ (they need not have a temporal order

with respect to some dynamics). The Pythagorean theorem of relative entropy states that, whenever the condition

$$\text{Tr}[(\theta - \omega)(\ln \omega - \ln \tau)] = 0 \quad (\text{B1})$$

is satisfied, it is possible to write the relative entropy between any pair of the three density matrices [e.g., $D(\theta \| \tau)$] as the sum of the relative entropies of the other two pairs [56], namely that

$$D(\theta \| \tau) = D(\theta \| \omega) + D(\omega \| \tau). \quad (\text{B2})$$

This is seen by a simple algebraic manipulation of the left-hand side, which generally leads to

$$\begin{aligned} D(\theta \| \tau) &= D(\theta \| \omega) + D(\omega \| \tau) \\ &\quad + \text{Tr}[(\theta - \omega)(\ln \omega - \ln \tau)]. \end{aligned} \quad (\text{B3})$$

However, when the condition in Eq. (B1) is not satisfied, we can still quantify the satisfiability of the Pythagorean theorem by using the trace norm as

$$|\text{Tr}[(\theta - \omega)(\ln \omega - \ln \tau)]| \leq \|(\theta - \omega)(\ln \omega - \ln \tau)\|_1, \quad (\text{B4})$$

which leads to the bound

$$\begin{aligned} |D(\theta \| \tau) - [D(\theta \| \omega) + D(\omega \| \tau)]| \\ \leq \|(\theta - \omega)(\ln \omega - \ln \tau)\|_1. \end{aligned} \quad (\text{B5})$$

Note that the right-hand side is still equal to zero when $(\theta - \omega)$ and $(\ln \omega - \ln \tau)$ satisfy the condition in Eq. (B1). Otherwise, using Hölder's inequality, we can bound the right-hand side as

$$\|(\theta - \omega)(\ln \omega - \ln \tau)\|_1 \leq \|\theta - \omega\|_1 \|\ln \omega - \ln \tau\|_\infty. \quad (\text{B6})$$

Therefore, we conclude our analysis by stating an approximate version of the Pythagorean theorem of relative entropy: if $\|\theta - \omega\|_1 \leq \varepsilon$ and $\|\ln \omega - \ln \tau\|_\infty \leq \delta$, then the Pythagorean theorem of relative entropy is satisfied with an error of at most $\varepsilon\delta$, namely

$$|D(\theta \| \tau) - [D(\theta \| \omega) + D(\omega \| \tau)]| \leq \varepsilon\delta. \quad (\text{B7})$$

-
- [1] E. T. Jaynes, Information theory and statistical mechanics, *Phys. Rev.* **106**, 620 (1957).
 - [2] V. C  pek and D. P. Sheehan, *Challenges to the Second law of Thermodynamics* (Springer, Dordrecht, The Netherlands, 2005).

- [3] K. Maruyama, F. Nori, and V. Vedral, Colloquium: The physics of Maxwell's demon and information, *Rev. Mod. Phys.* **81**, 1 (2009).
- [4] B. Groisman, S. Popescu, and A. Winter, Quantum, classical, and total amount of correlations in a quantum state, *Phys. Rev. A* **72**, 032317 (September 2005).
- [5] D. J. Bedingham and O. J. E. Maroney, The thermodynamic cost of quantum operations, *New J. Phys.* **18**, 113050 (2016).
- [6] K. Abdelkhalek, Y. Nakata, and D. Reeb, Fundamental energy cost for quantum measurement, [arXiv:1609.06981](https://arxiv.org/abs/1609.06981) (September 2016).
- [7] M. Berta, F. G. S. L. Brandão, C. Majenz, and M. M. Wilde, Conditional Decoupling of Quantum Information, *Phys. Rev. Lett.* **121**, 040504 (July 2018).
- [8] M. Berta, F. G. S. L. Brandão, C. Majenz, and M. M. Wilde, Deconstruction and conditional erasure of quantum correlations, *Phys. Rev. A* **98**, 042320 (October 2018).
- [9] S. Deffner, Energetic cost of Hamiltonian quantum gates, [arXiv:2102.05118](https://arxiv.org/abs/2102.05118) (February 2021).
- [10] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, *Phys. Rev. A* **52**, R2493 (1995).
- [11] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, United Kingdom, 2002).
- [12] D. Lidar and T. A. Brun, editors, *Quantum Error Correction* (Cambridge University Press, New York, NY, USA, 2013).
- [13] H. Barnum and E. Knill, Reversing quantum dynamics with near-optimal quantum and classical fidelity, *J. Math. Phys.* **43**, 2097 (May 2002).
- [14] H. Khoo Ng and P. Mandayam, Simple approach to approximate quantum error correction based on the transpose channel, *Phys. Rev. A* **81**, 062342 (June 2010).
- [15] J. Tyson, Two-sided bounds on minimum-error quantum measurement, on the reversibility of quantum dynamics, and on maximum overlap using directional iterates, *J. Math. Phys.* **51**, 092204 (September 2010).
- [16] M. Junge, R. Renner, D. Sutter, M. M. Wilde, and A. Winter, Universal recovery maps and approximate sufficiency of quantum relative entropy, *Annales Henri Poincaré* **19**, 2955 (2018).
- [17] H. Kwon, R. Mukherjee, and M. S. Kim, Reversing open quantum dynamics via continuous Petz recovery map, [arXiv:2104.03360](https://arxiv.org/abs/2104.03360) (2021).
- [18] V. Vedral, Landauer's erasure, error correction and entanglement, *Proc. R. Soc. London. Ser. A: Math. Phys. Eng. Sci.* **456**, 969 (2000).
- [19] V. Korepin and J. Terilla, Thermodynamic interpretation of the quantum error correcting criterion, *Quantum Inf. Process.* **1**, 225 (2002).
- [20] C. Cafaro and P. van Loock, An entropic analysis of approximate quantum error correction, *Phys. A: Stat. Mech. Appl.* **404**, 34 (2014).
- [21] H. J. Groenewold, A problem of information gain by quantal measurements, *Int. J. Theor. Phys.* **4**, 327 (1971).
- [22] M. Ozawa, On information gain by quantum measurements of continuous observables, *J. Math. Phys.* **27**, 759 (1986).
- [23] K. Jacobs, Second law of thermodynamics and quantum feedback control: Maxwell's demon with weak measurements, *Phys. Rev. A* **80**, 012322 (2009).
- [24] F. Buscemi, S. Das, and M. M. Wilde, Approximate reversibility in the context of entropy gain, information gain, and complete positivity, *Phys. Rev. A* **93**, 062314 (2016).
- [25] M. A. Nielsen, C. M. Caves, B. Schumacher, and H. Barnum, Information-theoretic approach to quantum error correction and reversible measurement, *Proc. R. Soc. London. Ser. A: Math. Phys. Eng. Sci.* **454**, 277 (1998).
- [26] G. T. Landi, A. L. Fonseca de Oliveira, and E. Buksman, Thermodynamic analysis of quantum error-correcting engines, *Phys. Rev. A* **101**, 042106 (2020).
- [27] T. Sagawa and M. Ueda, Second Law of Thermodynamics with Discrete Quantum Feedback Control, *Phys. Rev. Lett.* **100**, 080403 (2008).
- [28] A. E. Allahverdyan, R. Balian, and T. M. Nieuwenhuizen, Understanding quantum measurement from the solution of dynamical models, *Phys. Rep.* **525**, 1 (2013).
- [29] A. E. Allahverdyan, R. Balian, and T. M. Nieuwenhuizen, A sub-ensemble theory of ideal quantum measurement processes, *Ann. Phys.* **376**, 324 (2017).
- [30] F. Buscemi, M. Hayashi, and M. Horodecki, Global Information Balance in Quantum Measurements, *Phys. Rev. Lett.* **100**, 210504 (2008).
- [31] T. Albash, D. A. Lidar, M. Marvian, and P. Zanardi, Fluctuation theorems for quantum processes, *Phys. Rev. E* **88**, 032146 (2013).
- [32] J. Goold, M. Paternostro, and K. Modi, Nonequilibrium Quantum Landauer Principle, *Phys. Rev. Lett.* **114**, 060602 (2015).
- [33] F. Buscemi, D. Fujiwara, N. Mitsui, and M. Rotondo, Thermodynamic reverse bounds for general open quantum processes, *Phys. Rev. A* **102**, 032210 (2020).
- [34] S. Das, S. Khatr, G. Siopsis, and M. M. Wilde, Fundamental limits on quantum dynamics based on entropy change, *J. Math. Phys.* **59**, 012205 (2018).
- [35] M. Ozawa, Quantum measuring processes of continuous observables, *J. Math. Phys.* **25**, 79 (1984).
- [36] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, United Kingdom, February 2017), 2nd ed.
- [37] M. Hayashi, *Quantum Information* (Springer, Tokyo, Japan, 2006).
- [38] A. Winter, "Extrinsic" and "intrinsic" data in quantum measurements: Asymptotic convex decomposition of positive operator valued measures, *Commun. Math. Phys.* **244**, 157 (January 2004).
- [39] M. M. Wilde, P. Hayden, F. Buscemi, and M.-H. Hsieh, The information-theoretic costs of simulating quantum measurements, *J. Phys. A: Math. Theor.* **45**, 453001 (2012).
- [40] M. Berta, J. M. Renes, and M. M. Wilde, Identifying the information gain of a quantum measurement, *IEEE Trans. Inf. Theory* **60**, 7987 (December 2014).
- [41] F. Leditzky, M. M. Wilde, and N. Datta, Strong converse theorems using Rényi entropies, *J. Math. Phys.* **57**, 082202 (2016).
- [42] A. Anshu, R. Jain, and N. A. Warsi, Convex-split and hypothesis testing approach to one-shot quantum measurement compression and randomness extraction, *IEEE Trans. Inf. Theory* **65**, 5905 (2019).

- [43] T. A. Atif, M. Heidari, and S. S. Pradhan, Faithful simulation of distributed quantum measurements with applications in distributed rate-distortion theory. *To appear in IEEE Transactions on Information Theory*, [arXiv:1901.06788](#) (2019).
- [44] I. Devetak and A. Winter, Classical data compression with quantum side information, *Phys. Rev. A* **68**, 042301 (October 2003).
- [45] L. del Rio, J. Åberg, R. Renner, O. Dahlsten, and V. Vedral, The thermodynamic meaning of negative entropy, *Nature* **474**, 61 (2011).
- [46] B. Schumacher and M. A. Nielsen, Quantum data processing and error correction, *Phys. Rev. A* **54**, 2629 (1996).
- [47] B. Schumacher, Sending entanglement through noisy quantum channels, *Phys. Rev. A* **54**, 2614 (1996).
- [48] E. Knill and R. Laflamme, Theory of quantum error-correcting codes, *Phys. Rev. A* **55**, 900 (1997).
- [49] T. Ogawa, Perfect quantum error-correcting condition revisited, [arXiv:quant-ph/0505167](#) (May 2005).
- [50] For example, we can consider the thermal state $\tau^B = \frac{1}{Z} \sum_{\alpha} e^{-\beta \epsilon_{\alpha}} |\alpha\rangle\langle\alpha|^B$ of a thermal bath as part of a larger environment $E = BB \cdots B$, comprised of N identical subsystems, in the pure state $|\beta\rangle^E = \frac{1}{\sqrt{Z}} \sum_{\alpha} e^{-\frac{\beta \epsilon_{\alpha}}{2}} |\alpha\rangle^{\otimes N}$, hence $\tau^B = \text{Tr}_{E \setminus B}[|\beta\rangle\langle\beta|^E]$. Consequently, a quantum process of the form $\rho_i^A \rightarrow \rho_f^A = \text{Tr}_B[U(\rho_i^A \otimes \rho_i^B)U^\dagger]$ can be written in the Stinespring form $\rho_f^A = \text{Tr}_E[(U \otimes V^{E \setminus B})(\rho_i^A \otimes |\beta\rangle\langle\beta|^E)(U \otimes V^{E \setminus B})^\dagger]$, where $V^{E \setminus B}$ is an arbitrary unitary on the extended environment $E \setminus B$.
- [51] R. Balian, *From Microphysics to Macrophysics: Methods and Applications of Statistical Physics* (Springer Science & Business Media, Heidelberg, Germany, 2007), Vol. II.
- [52] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, New York, NY, USA, 2002).
- [53] M. Esposito, K. Lindenberg, and C. Van den Broeck, Entropy production as correlation between system and reservoir, *New J. Phys.* **12**, 013013 (2010).
- [54] C. Kwon, J. Um, J. Yeo, and H. Park, Three heats in a strongly coupled system and bath, *Phys. Rev. E* **100**, 052127 (2019).
- [55] D. Reeb and M. M. Wolf, An improved Landauer principle with finite-size corrections, *New J. Phys.* **16**, 103011 (2014).
- [56] S. Weis, Information topologies on non-commutative state spaces, [arXiv:1003.5671](#) (2013).
- [57] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and Th. Beth, Thermodynamic cost of reliability and low temperatures: Tightening Landauer's principle and the second law, *Int. J. Theor. Phys.* **39**, 2717 (2000).
- [58] M. Horodecki, P. W. Shor, and M. B. Ruskai, Entanglement breaking channels, *Rev. Math. Phys.* **15**, 629 (2003).
- [59] P. Faist, F. Dupuis, J. Oppenheim, and R. Renner, The minimal work cost of information processing, *Nat. Commun.* **6**, 1 (2015).
- [60] M. Hayashi and H. Tajima, Measurement-based formulation of quantum heat engines, *Phys. Rev. A* **95**, 032132 (2017).
- [61] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, Experimental verification of Landauer's principle linking information and thermodynamics, *Nature* **483**, 187 (2012).
- [62] N. Zhao, and S.-W. Ho, R.-B. Liu, Decoherence and dynamical decoupling control of nitrogen vacancy center electron spins in nuclear spin baths, *Phys. Rev. B* **85**, 115303 (2012).
- [63] Z.-H. Wang, G. De Lange, D. Ristè, R. Hanson, and V. V. Dobrovitski, Comparison of dynamical decoupling protocols for a nitrogen-vacancy center in diamond, *Phys. Rev. B* **85**, 155204 (2012).
- [64] T. Sagawa and M. Ueda, Minimal Energy Cost for Thermodynamic Information Processing: Measurement and Information Erasure, *Phys. Rev. Lett.* **102**, 250602 (2009).
- [65] K. Funo, Y. Watanabe, and M. Ueda, Thermodynamic work gain from entanglement, *Phys. Rev. A* **88**, 052319 (2013).
- [66] J. Jun Park, K.-H. Kim, T. Sagawa, and S. W. Kim, Heat Engine Driven by Purely Quantum Information, *Phys. Rev. Lett.* **111**, 230402 (2013).
- [67] A. C. Barato and U. Seifert, Thermodynamic Uncertainty Relation for Biomolecular Processes, *Phys. Rev. Lett.* **114**, 158101 (2015).
- [68] P. Pietzonka, A. C. Barato, and U. Seifert, Universal bound on the efficiency of molecular motors, *J. Stat. Mech.: Theory Exp.* **2016**, 124004 (2016).
- [69] P. Pietzonka and U. Seifert, Universal Trade-Off between Power, Efficiency, and Constancy in Steady-State Heat Engines, *Phys. Rev. Lett.* **120**, 190602 (2018).
- [70] H. J. D. Miller, M. Hamed Mohammady, M. Perarnau-Llobet, and G. Guarnieri, Thermodynamic Uncertainty Relation in Slowly Driven Quantum Heat Engines, *Phys. Rev. Lett.* **126**, 210603 (2021).
- [71] A. Kay, Tutorial on the QUANTIKZ package, [arXiv preprint arXiv:1809.03842](#) (2018).
- [72] M. Perarnau-Llobet, K. V. Hovhannisyan, M. Huber, P. Skrzypczyk, N. Brunner, and A. Acín, Extractable Work from Correlations, *Phys. Rev. X* **5**, 041011 (2015).
- [73] G. A. Paz-Silva, M. J. W. Hall, and H. M. Wiseman, Dynamics of initially correlated open quantum systems: Theory and applications, *Phys. Rev. A* **100**, 042120 (2019).
- [74] A. E. Allahverdyan and T. M. Nieuwenhuizen, Breakdown of the Landauer bound for information erasure in the quantum regime, *Phys. Rev. E* **64**, 056117 (2001).
- [75] A. E. Allahverdyan, R. Balian, and T. M. Nieuwenhuizen, in *AIP Conference Proceedings*, (American Institute of Physics, 2002), Vol. 643, p. 35.