Work extraction and fully entangled fraction

Chung-Yun Hsieh[1,*] and Ray-Kuang Lee[1,2]

[1]*Department of Physics, National Tsing Hua University, Hsinchu 300, Taiwan*
[2]*Physics Division, National Center for Theoretical Sciences, Hsinchu 300, Taiwan*
(Received 9 June 2017; published 10 July 2017)

For a bipartite state with equal local dimension d, we prove that one can obtain work gain under Landauer's erasure process on one party in the identically and independently distributed limit when the corresponding fully entangled fraction is larger than $\frac{1}{d}$. By processing a given state to the maximally mixed state, we give an approximation for the largest extractable work with an error in the energy scale, which becomes negligible in the large system limit. As a step to link quantum thermodynamics and quantum nonlocality, we also provide a simple picture to approximate the optimal work extraction and suggest a potential thermodynamic interpretation of the fully entangled fraction for isotropic states.

I. INTRODUCTION

Quantum thermodynamics and quantum nonlocality share similar capacity to tell quantum and classical regimes apart. For quantum nonlocality, we have the famous Einstein-Podolsky-Rosen (EPR) paradox [1] and Bell's inequality [2] to illustrate the bizarre nature of quantum theory. On the other hand, for quantum thermodynamics, a multitude of intriguing phenomena related to various definitions of *work* have been addressed [3]. Examples such as Landauer's principle [4,5], different scenarios on work extraction [6–8], the minimal work cost of running different processes such as a completely positive trace-preserving map (CPTPM) [9,10], or an erasure process [11], and other related results [12–18] certify the importance in this direction.

Even though the awareness of the relations between work and entanglement [13], coherence [19,20], or other quantum correlations [14,16,21] has already been studied in the literature, a quantitative connection to nonlocal properties such as quantum nonlocality [22], quantum steerability [23,24], and the usefulness of teleportation [25–28] still remains as an open question deserving further study. To this end, we adapt a well-known quantity called *fully entangled fraction* (FEF) [27,28], which acts as a useful measurement in characterizing different nonlocal properties [22,29–32]. In this work, we try to bridge quantum thermodynamics and quantum nonlocality together, by relating work gain under different processes to FEF.

With the help of FEF, for a given bipartite quantum state ρ acting on $\mathbb{C}^d \otimes \mathbb{C}^d$, we derive an inequality to conclude that in the identically and independently distributed (iid) limit, it is always possible to have work gain (i.e., negative work cost) of an erasure process on one party of ρ, provided that ρ has a FEF larger than $\frac{1}{d}$. This result gives an alternative interpretation for every state useful for teleportation [25,27], which is equivalent to enabling work gain under the local erasure process. For a specified temperature T, we define work extraction to be a process mapping the initial state with a fully degenerate system Hamiltonian to the final Gibbs state in T with the Hamiltonian equal to the initial one. With some prerequisites, we obtain an approximation for the largest extractable work, up to an error in

the energy scale which can be dropped when the system is large enough. This approximation not only provides a clear strategy for work extraction, which consists of two local operations and classical communications (LOCCs) plus one local erasure process, but also suggests we interpret the FEF for isotropic states [27] as a concept equivalent to the minimal work cost of the erasure process on one party.

This paper is structured as follows. The notations used and the underline scenario to define the erasure process and the work extraction are described in Sec. II. Then, in Sec. III, the first main result (Theorem 1) is illustrated and proved, which gives an inequality linking the FEF and the work gain of the local erasure process. The saturation for this inequality is discussed in Sec. IV, as a corollary of the approximation of the optimal extractable work (Theorem 2). Furthermore, we apply this result to isotropic states and arbitrary pure states in the remaining part of Sec. IV, where the former suggests to us a thermodynamic interpretation of the FEF and the latter suggests a nearly local protocol to extract the maximal work of an arbitrary pure state. Finally, in Sec. V we give the conclusion.

II. PRELIMINARY NOTIONS

Throughout this paper, we only consider quantum states ρ acting on a bipartite system, $\mathbb{C}^d \otimes \mathbb{C}^d$, with equal local dimension $d = 2^l$ for some natural number $l \in \mathbb{N}$. For such a system, we call the first party "Alice" and the second party "Bob." The collection of all quantum states acting on this bipartite system is denoted as $\mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$. The *conditional von Neumann entropies* (see, for example, Ref. [33]) of a given state $\rho \in \mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$ are

$$S(A|B)_\rho := S(\rho) - S(\rho_B),$$
$$S(B|A)_\rho := S(\rho) - S(\rho_A), \tag{1}$$

where $S(\rho) := -\text{tr}(\rho \log_2 \rho)$ is the *von Neumann entropy* of the quantum state ρ and $\rho_A := \text{tr}_B(\rho)$ [$\rho_B := \text{tr}_A(\rho)$] is the local state on Alice's (Bob's) side.

The FEF [27,28] of a given state $\rho \in \mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$ is defined by

$$\mathcal{F}(\rho) := \max_{|\Psi\rangle} \langle \Psi | \rho | \Psi \rangle, \tag{2}$$

*s103022502@m103.nthu.edu.tw

where the maximization is taken over all maximally entangled states $|\Psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$. It is well known that the FEF has been successfully applied to characterize different nonlocal properties. For instance, a quantum state $\rho \in \mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$ is useful for the standard teleportation if and only if $\mathcal{F}(\rho) > \frac{1}{d}$ [27]; while ρ is k-copy nonlocal (and hence k-copy steerable) if $\mathcal{F}(\rho) > \frac{1}{d}$ [30]. Sufficient conditions for quantum steerability can also be given by the FEF [31,32]. Therefore, if one is able to link the FEF with quantum thermodynamics, it can be a good starting point for bridging quantum nonlocality and quantum thermodynamics.

For a given state $\rho \in \mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$, we want to consider extractable work from the *information content* of ρ. More precisely, we tend to define work extraction by asking the following: *how much work can be extracted from ρ by processing it to the maximally mixed state $\frac{\mathbb{I}}{d^2}$?* The reason to adapt this definition is because we expect the maximally mixed state to contain the *least* useful information content. Therefore processing ρ to it should be a natural way to obtain the optimal amount of work from the useful information possessed by ρ.

To study work extraction and other processes, rigorously, we need to specify our framework, which is inspired by Ref. [11]. Let us consider that the system $\mathbf{S} = \mathbb{C}^d \otimes \mathbb{C}^d$ couples to an environment $\mathbf{E}$, so that $\mathbf{S} \otimes \mathbf{E}$ is a closed system with only unitary time evolution. We require the initial system Hamiltonian to be fully degenerate. This means any computational basis will be an energy eigenbasis; in particular, we are allowed to choose the computational basis in the form $\{|n_1\rangle \otimes |n_2\rangle \otimes \cdots \otimes |n_l\rangle \,|\, n_1, n_2, \dots, n_l \in \{0,1\}\}$, where $\{|0\rangle, |1\rangle\}$ is a single qubit computational basis. The environment consists of two parts: a heat bath $\mathbf{B}_T$, in a finite and positive temperature T, and an observer $\mathbf{O}$. We further assume the observer possesses a subsystem, called a *battery*, which is assumed to be the perfect energy storage to store or withdraw energy without loss. The battery is also assumed to be able to generate energy in the form of work. For example, it can be a collection of work qubits [6] or a suspended weight that will be raised or lowered [7,12]. In this sense, if energy is extracted from the system by the observer and stored in the battery, one can output the same amount of work from the battery. Theoretically and ideally, it is in this sense that we talk about "work."

With the above setting, we can now consider the process that may help us to draw work from the system. For this purpose, we introduce the *allowed actions* for the observer $\mathbf{O}$ to apply, which amounts to the observer's abilities to manipulate the system.

(a) Raising or lowering the energy level. This action allows the observer to raise or lower the energy level of the system (e.g., by adjusting the magnetic field), which is a crucial action for work extraction. If the system is at an energy level n and the observer changes the corresponding energy from E_n to E'_n, we define the *work gain* (for the observer) to be $E'_n - E_n$ [11]. Note that we call it work gain based on the assumption of the existence of the ideal battery. This definition can be generalized to arbitrary states $\rho \in \mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$ by $\mathrm{tr}[\rho(H' - H)]$, where H is the initial system Hamiltonian and H' is the final Hamiltonian after this action, i.e., raising or lowering manipulation. We use $\mathcal{A}_{\mathrm{R/L}}$ to denote the set of all actions of this type.

(b) Thermalization. Within this action, the observer will make (part of) the system couple to the heat bath $\mathbf{B}_T$ to achieve thermalization, and then the system will decouple from $\mathbf{B}_T$ at the end of this action. This amounts to running some CPTPMs on the system, which is mainly used to make the system go to the Gibbs state. Although a CPTPM does need work cost [10], we assume that this cost is covered by the heat bath rather than the observer. Hence, we assume no energy change of the observer in this action. We use $\mathcal{A}_T$ to denote the set of all actions of this type with temperature T.

(c) Unitary operations. This action includes mappings in the form $\rho \mapsto \int_{U(d^2)} U \rho U^\dagger P_U \, dU$ for some probability density function P_U over the unitary group $U(d^2)$. Also, we allow the observer to select specific results and abandon the rest, resulting in nonzero failure probability. Note that unlike the actions of thermalization, this action needs work cost from the observer [10]. In the case of negative work cost, we define it to be work gain (upon success) given the existence of the ideal battery. We use $\mathcal{A}_U$ to denote the set of all actions of this type.

Apart from the actions of unitary operations, we also allow the following source of probabilistic nature, which is a special action for the observer:

(d) δ approximation. Within this action, the observer is allowed to approximate the original input state by another state which is δ *closed* to it (a state ρ is δ *closed* to another state σ if their trace distance [9] is upper bounded by δ, i.e., $\frac{1}{2}\|\rho - \sigma\|_1 \leqslant \delta$). This *conceptual* action has no real effect on the state, and it is introduced so that in our framework one can define allowed processes rigorously and completely. Note that this step will result in nonzero failure probability. We use $\mathcal{A}_\delta$ to denote the set of all actions with δ precision of this type.

Now, we are in a position to define process in our framework: for a given state $\rho \in \mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$, we define an *allowed process* for ρ, denoted by $\mathcal{P}_\rho$, to be a finite sequence of actions, $\mathcal{P}_\rho = \{A_n\}_{n=1}^N \subset \mathcal{A}_{\mathrm{R/L}} \cup \mathcal{A}_T \cup \mathcal{A}_U \cup \mathcal{A}_\delta$, such that its behavior on the system can be mathematically written as $A_N \circ A_{N-1} \circ \cdots \circ A_1$, where A_1 is acting on ρ with the initial system Hamiltonian H (fully degenerate) and A_N will output a final state with the final system Hamiltonian being equal to the original one, i.e., H. Note that due to this definition, the observer can apply one and only one action at one time. To illustrate our idea, let us consider the proof of Theorem I.1 in del Rio *et al.* [11] as an example. In their proof, in order to construct an erasure process on the given state ρ, they first choose a subsystem which is δ closed to a maximally entangled pure state $|\Psi\rangle$, and then they apply work extraction "for $|\Psi\rangle$" on the given system. The work extraction they use for $|\Psi\rangle$ consists of actions in $\mathcal{A}_{R/L} \cup \mathcal{A}_T$ only (see Fig. 4 in Ref. [11]). In our framework, such a process can therefore be written as $\mathcal{P}_\rho = \{A_n\}_{n=1}^N$, where $A_1 \in \mathcal{A}_\delta$ and $\{A_n\}_{n=2}^N \subset \mathcal{A}_{R/L} \cup \mathcal{A}_T$. As mentioned previously, the action of δ approximation results in a probabilistic nature. This can be seen by the failure probability for the above process, which is less than δ as proved by Supplementary Lemma IV.3 in Ref. [11].

Now, we define $W_A(\rho)$ to be the work gain for an action A on ρ upon success. Following this definition, we can define the *work gain* of an allowed process $\mathcal{P}_\rho = \{A_n\}_{n=1}^N$, denoted by $W(\mathcal{P}_\rho)$, to be $W(\mathcal{P}_\rho) := \sum_{n=1}^N W_{A_n}(\rho)$, upon the success of all the involved actions. This also explains why we define an allowed process to be *a chain of actions* rather than a single

mapping: if one defines a process to be a mapping, it may be ill-defined physically: there may be different chains of actions with *different net work gains* resulting in the same mapping, which is, however, not the physical property that we expect.

III. LOCAL ERASURE PROCESS AND FULLY ENTANGLED FRACTION

Consider a given quantum state $\rho \in \mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$. Within our framework introduced in Sec. II, we say $\mathcal{P}_\rho^{\mathrm{Er|A}}$ is a *local (Landauer's) erasure process* for ρ on Alice's side if (i) $\mathcal{P}_\rho^{\mathrm{Er|A}} = \mathcal{P}_{|\phi\rangle\langle\phi|}$, where $|\phi\rangle\langle\phi|$ is a purification of ρ (and the additional space needed for the purification is assumed to be included in the space of the observer), and (ii) $\mathcal{P}_{|\phi\rangle\langle\phi|}$ maps $|\phi\rangle\langle\phi| \mapsto |\mathbf{0}\rangle\langle\mathbf{0}| \otimes \mathrm{tr}_{\mathrm{A}}|\phi\rangle\langle\phi|$, i.e., the state outside Alice's local system will be preserved. Here, we use the notation $|\mathbf{0}\rangle\langle\mathbf{0}| := (|0\rangle\langle 0|)^{\otimes l}$. This shows why we call it a *local* erasure process. Also, we only consider the *minimal deterministic average work cost* in the iid limit of erasure process defined as follows [11] (note that we have a minus sign here because we always reserve the positive value for work gain):

$$
\begin{aligned}
- W_{\mathrm{Er}}(\rho) &:= \inf \big\{ w \big| \exists \big\{ \mathcal{P}_{\rho^{\otimes k}}^{\mathrm{Er|A}} \big\}_{k=1}^\infty \text{ s.t.} \\
&\quad \lim_{k\to\infty} P\big[-W\big(\mathcal{P}_{\rho^{\otimes k}}^{\mathrm{Er|A}}\big) \leqslant kw \big] = 1 \big\},
\end{aligned} \tag{3}
$$

where $P[-W(\mathcal{P}) \leqslant x]$ is the probability for the process $\mathcal{P}$ to achieve work cost $-W(\mathcal{P})$ no greater than x (recall the probabilistic nature of allowed actions). To understand the meaning of this definition, note that only the regime $k \gg 1$ will be important. Hence, $-W_{\mathrm{Er}}(\rho)k$ is the minimal work cost that can be achieved almost with certainty for any large enough k, and the probability of failure will go to zero when $k \to \infty$. This justifies the interpretation of minimal *deterministic* average work cost in the iid limit. From Ref. [11], we have the following upper bounds:

$$
- W_{\mathrm{Er}}(\rho) \leqslant S(A|B)_\rho k_B T \ln 2. \tag{4}
$$

The remarkable consequence of the above inequality is the existence of the negative work cost of the local erasure process for some entangled states [11], which also implies that $W_{\mathrm{Er}}(\rho) > 0$ serves as an entanglement witness. However, how strong the entanglement needs to be in order to have work gain under the local erasure process is still unknown. This is one example of questions that can be answered by our first main result, which is a direct consequence of the following lemma, whose proof can be found in Appendix A:

Lemma 1. If $\mathcal{F}(\rho) > \frac{1}{d}$, then

$$
S(A|B)_\rho \leqslant - \log_2 \mathcal{F}(\rho)d. \tag{5}
$$

Combining Eq. (4) and Lemma 1, we obtain our first main result:

Theorem 1. If $\mathcal{F}(\rho) > \frac{1}{d}$, then

$$
W_{\mathrm{Er}}(\rho) \geqslant k_B T \ln \mathcal{F}(\rho)d. \tag{6}
$$

In other words, states with a FEF larger than $\frac{1}{d}$ enable work gain under the local erasure process.

A. Theorem 1 and teleportation

Several remarks can be made accordingly. As the first one, we recall the famous result developed by Horodecki *et al.* [27] that a given state $\rho \in \mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$ is useful for teleportation if and only if $\mathcal{F}(\rho) > \frac{1}{d}$. Together with Theorem 1, one learns that *a state $\rho \in \mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$ admits deterministic work gain in the iid limit under the local erasure process if it is useful for teleportation.* This build a simple but quantitative link between the erasure process and the teleportation power. Also, Ref. [11] shows that $W_{\mathrm{Er}}(\rho) > 0$ is a *sufficient* condition of entanglement. Via Theorem 1, one is able to partially address the converse—we learn that $\mathcal{F}(\rho) > \frac{1}{d}$ is a sufficient condition of $W_{\mathrm{Er}}(\rho) > 0$. This means when the entanglement is strong enough (in the sense that FEF $> \frac{1}{d}$), there exists a local erasure process inducing work gain.

B. Theorem 1 and nonlocal properties

As another remark, Theorem 1 can be used to establish a sufficient condition of locality and nonsteerability for a certain class of quantum states with FEF $> \frac{1}{d}$. To see this, we note that Theorem 1 implies the following upper bound on the FEF of the given state ρ with $\mathcal{F}(\rho) > \frac{1}{d}$:

$$
\mathcal{F}(\rho) \leqslant \frac{1}{d} e^{\frac{W_{\mathrm{Er}}(\rho)}{k_B T}}. \tag{7}
$$

As an example, consider the isotropic state [27] defined by

$$
\rho_{\mathrm{iso}}(p) := p|\Psi_d^+\rangle\langle\Psi_d^+| + (1-p)\frac{\mathbb{I}}{d^2}, \tag{8}
$$

where $|\Psi_d^+\rangle := \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |ii\rangle$ is the generalized singlet and $p \in [-\frac{1}{d^2-1}, 1]$ due to the positivity of a quantum state. As a direct corollary, since ρ_{iso} is entangled if and only if $\mathcal{F}(\rho_{\mathrm{iso}}) > \frac{1}{d}$ [22], Theorem 1 implies ρ_{iso} *admits work gain under the local erasure process if and only if ρ_{iso} is entangled.* For other nonlocal correlations, we make use of the well-known property of isotropic states: there exist many thresholds for different nonlocal properties [22]. For instance, there exist $\mathcal{F}_{\mathrm{LHV}}$, $\mathcal{F}_{\mathrm{LHS}}^\pi$, and $\mathcal{F}_{\mathrm{LHS}}$ such that ρ_{iso} is local under general positive operator-valued measures (POVMs) if $\mathcal{F}(\rho_{\mathrm{iso}}) \leqslant \mathcal{F}_{\mathrm{LHV}}$ [22,34]; ρ_{iso} is unsteerable under projective POVMs if and only if $\mathcal{F}(\rho_{\mathrm{iso}}) \leqslant \mathcal{F}_{\mathrm{LHS}}^\pi = \frac{H_d + H_d d - d}{d^2}$, where $H_d := \sum_{n=1}^d \frac{1}{n}$ [24], and ρ_{iso} is unsteerable under general POVMs if $\mathcal{F}(\rho_{\mathrm{iso}}) \leqslant \mathcal{F}_{\mathrm{LHS}} = \tilde{p}_\phi(1 - \frac{1}{d^2}) + \frac{1}{d^2}$, where $\tilde{p}_\phi := \frac{3d-1}{d^2-1}(1 - \frac{1}{d})^d$ [34]. Applying Eq. (7) on ρ_{iso} with FEF $> \frac{1}{d}$, we learn that it is local if

$$
W_{\mathrm{Er}}(\rho_{\mathrm{iso}}) \leqslant k_B T \ln \mathcal{F}_{\mathrm{LHV}} d; \tag{9}
$$

it is unsteerable under general POVMs if

$$
W_{\mathrm{Er}}(\rho_{\mathrm{iso}}) \leqslant k_B T \ln \left[\tilde{p}_\phi\left(d - \frac{1}{d}\right) + \frac{1}{d} \right]; \tag{10}
$$

it is unsteerable under projective POVMs if

$$
W_{\mathrm{Er}}(\rho_{\mathrm{iso}}) \leqslant k_B T \ln \frac{H_d + H_d d - d}{d}. \tag{11}
$$

Because there is a hierarchy consisting of different thresholds for the FEF of isotropic states, one may wonder whether Eq. (7) can map this hierarchy onto the one consisting of

different erasure work costs. This can be achieved if the upper bound in Eq. (7) is saturated by isotropic states. We come back to this issue in Sec. IV A.

IV. WORK EXTRACTION PROCESS AND FULLY ENTANGLED FRACTION

In this section, we want to study the relation between work extraction and the fully entangled fraction. Before proceeding, let us first define the former. Consider a given state $\rho \in \mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$. We say $\mathcal{P}_\rho^{\mathrm{W}}$ is a *work extraction process* for ρ if it is an allowed process having $\frac{\mathbb{I}}{d^2}$ as the final state. Note that the final Hamiltonian is also fully degenerate in our formalism, which means that the maximally mixed state equals the Gibbs state in the given temperature. Due to this definition, we define the *largest deterministic extractable work* in the iid limit for ρ as

$$W_{\mathrm{Total}}(\rho) := \sup \left\{ w \mid \exists \{ \mathcal{P}_{\rho^{\otimes k}}^{\mathrm{W}} \}_{k=1}^\infty \text{ s.t.} \right.$$
$$\left. \lim_{k \to 1} P\left[W\left(\mathcal{P}_{\rho^{\otimes k}}^{\mathrm{W}} \right) \geqslant kw \right] = 1 \right\}. \quad (12)$$

Just like the definition of erasure work cost, $W_{\mathrm{Total}}(\rho)$ serves as the maximal work gain from ρ with certainty in the iid limit. This definition should be an answer to our initial question in the iid limit, i.e., how much work can be extracted by processing ρ to $\frac{\mathbb{I}}{d^2}$?

As a direct observation from Eq. (4), since the composition of processes $\mathcal{P}_{|\mathbf{0}\rangle\langle\mathbf{0}|^{\otimes k}}^{\mathrm{W}} \circ \mathcal{P}_{\rho^{\otimes k}}^{\mathrm{Er}}$ is a work extraction process, we conclude that (see Appendix C for the proof)

$$W_{\mathrm{Total}}(\rho) \geqslant k_B T \ln d^2 - S(\rho) k_B T \ln 2. \quad (13)$$

Before stating the main result, we need to introduce an estimate done by Dahlsten *et al.* [12]. Define an ϵ-*compression* action for ρ to be an action of unitary operations mapping as $\rho \mapsto U\rho U^\dagger$ such that $U\rho U^\dagger$ is 2ϵ close to a state of the form $\rho' \otimes |\psi\rangle\langle\psi|$, with success probability $P_{\mathrm{success}} \geqslant 1 - 2\epsilon$ [35]. Then the process adapted by Ref. [12] is given by $\{A_n\}_{n=1}^N$, where A_1 is an ϵ-compression action and $\{A_n\}_{n=2}^N$ amounts to a work extraction process on $|\psi\rangle\langle\psi|$ with work gain $k_B T \ln d_\psi$, where d_ψ is the dimension of the local system $|\psi\rangle\langle\psi|$. We call them *compression-extraction* processes and use the notation $\mathcal{P}_\rho^\epsilon$ to denote such processes with $P_{\mathrm{success}} \geqslant 1 - 2\epsilon$. Then Theorem 2 in Ref. [12] states that if the extractable work by a compression-extraction process of a given bipartite state $\rho \in \mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$ is lower bounded by $k_B T \ln d^2 - [H_{\min}^\epsilon(\rho) + 3 \ln \epsilon] k_B T \ln 2$, then we have $P_{\mathrm{success}} < 2\epsilon$. Here $H_{\min}^\epsilon(\rho)$ is the smooth min-entropy of ρ, whose definition can be found in Eq. (D3) (we refer the readers to Refs. [12,33] and references therein for the detail). This means whenever one is able to choose some $\mathcal{P}_\rho^{\frac{1}{2}-\epsilon}$ to extract $W_{\mathrm{Total}}(\rho)$ (therefore with $P_{\mathrm{success}} \geqslant 2\epsilon$), Theorem 2 in Ref. [12] becomes an *upper* bound on it: $W_{\mathrm{Total}}(\rho) < k_B T \ln d^2 - [H_{\min}^\epsilon(\rho) + 3 \ln \epsilon] k_B T \ln 2$. With Lemma 1 in hand, we are now able to estimate $W_{\mathrm{Total}}(\rho)$ in terms of FEF and conditional von Neumann entropy under the above conditions. [We refer the reader to Appendix D for the proof; also, we adapt the notation $S_{\min}(\rho) := \min\{S(\rho_A); S(\rho_B)\}$.]

Theorem 2. Given $0 < \epsilon \leqslant \frac{1}{2}$ and $\delta_\epsilon := -3 \ln \epsilon$, if (i) $\mathcal{F}(\rho) > \frac{1}{d}$, (ii) there exists $\mathcal{P}_\rho^{\frac{1}{2}-\epsilon}$ which can extract $W_{\mathrm{Total}}(\rho)$,
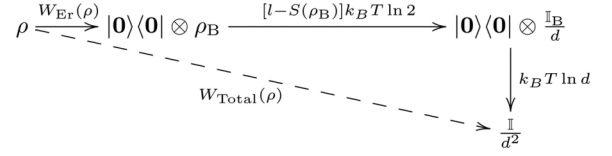

FIG. 1. Schematic interpretation of Eq. (14). Assume $S_{\min}(\rho) = S(\rho_B)$ without a loss of generality. In this diagram, via an optimal local erasure process in the iid limit on Alice's side (the upper-left arrow), a work extraction on Bob's side (the upper-right arrow), and a work extraction on Alice's side (the downward arrow), $W_{\mathrm{Total}}(\rho)$ is extracted (the dashed arrow), up to $\delta_\epsilon k_B T \ln 2$.

and (iii) $\log_2 \|\rho\|_\infty = \log_2 \mathcal{F}(\rho)d - S_{\min}(\rho)$, then

$$W_{\mathrm{Total}}(\rho) \approx k_B T \ln d^2 - S_{\min}(\rho) k_B T \ln 2 + W_{\mathrm{Er}}(\rho) \quad (14)$$

and

$$W_{\mathrm{Er}}(\rho) \approx k_B T \ln \mathcal{F}(\rho)d, \quad (15)$$

up to an error of $\delta_\epsilon k_B T \ln 2$.

As a remark, Eq. (14) provides a picture (see Fig. 1) of an approximately optimal *global* deterministic work extraction in the iid limit. Note that the above approximation will be faithful if $\delta_\epsilon k_B T \ln 2 \ll k_B T \ln d$, i.e., if $\delta_\epsilon \ll l$, where l is the number of qubits in a single party ($d = 2^l$). In other words, $\delta_\epsilon k_B T \ln 2$, which is the best resolution for energy in this case, will be an irrelevant scale. Theorem 2 can be regarded as an approximately sufficient condition of the tightness of Theorem 1, and thus Eq. (4) and Lemma 1. In particular, approximate saturation for Lemma 1 together with condition (3) in the above theorem implies the following corollary:

$$S(\rho) \approx -\log_2 \|\rho\|_\infty, \quad (16)$$

up to an error of δ_ϵ.

To illustrate the applications of Theorem 2, let us compute some examples. In particular, what we want is a thermodynamic interpretation of the FEF for certain quantum states.

Due to later consideration, let us define $\Lambda(\rho) := \log_2 \|\rho\|_\infty - [\log_2 \mathcal{F}(\rho)d - S_{\min}(\rho)]$. The following subsections hold for ϵ values satisfying Theorem 2.

A. Theorem 2 and isotropic states

From Eq. (8), one can see that $\Lambda[\rho_{\mathrm{iso}}(p)] = 0$ *for all p* [to prove this, it suffices to note that $\mathcal{F}(\rho_{\mathrm{iso}}) = \|\rho_{\mathrm{iso}}\|_\infty$ and $S_{\min}(\rho_{\mathrm{iso}}) = \log_2 d$]. This means for isotropic states with FEF $> \frac{1}{d}$, if there exists $\mathcal{P}_{\rho_{\mathrm{iso}}}^{\frac{1}{2}-\epsilon}$ which can extract $W_{\mathrm{Total}}(\rho_{\mathrm{iso}})$, then $W_{\mathrm{Total}}(\rho_{\mathrm{iso}}) \approx k_B T \ln d + W_{\mathrm{Er}}(\rho_{\mathrm{iso}})$ and $W_{\mathrm{Er}}(\rho_{\mathrm{iso}}) \approx k_B T \ln \mathcal{F}(\rho_{\mathrm{iso}})d$ up to an energy scale of $\delta_\epsilon k_B T \ln 2$. In particular, the latter suggests a possible thermodynamic interpretation of the FEF for isotropic states—*up to $\delta_\epsilon k_B T \ln 2$; the FEF of isotropic states is a concept equivalent to the minimal deterministic work cost in the iid limit of the local erasure process.*

Also note that $W_{\mathrm{Er}}(\rho_{\mathrm{iso}}) \approx k_B T \ln \mathcal{F}(\rho_{\mathrm{iso}})d$ builds an approximate hierarchy according to the result discussed in Sec. III B from the approximate saturation of Eq. (7); namely, by substituting different FEF thresholds of nonlocal properties for isotropic states into the saturation bound, one obtains the corresponding erasure work cost threshold for isotropic states,
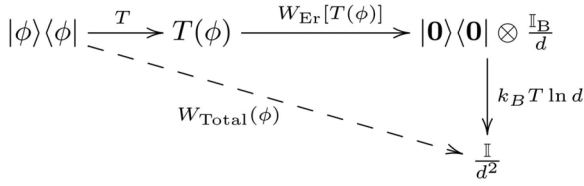
FIG. 2. Schematic interpretation for the optimal deterministic work extraction in the iid limit up to $\delta_\epsilon k_B T \ln 2$. The upper-left arrow is the quantum twirling bringing the input pure state to an isotropic state, and the remaining processes, i.e., the upper-right and downward arrows, are just the consequence of Eq. (14) on isotropic states.

up to the energy scale $\delta_\epsilon k_B T \ln 2$. For example, if ρ_{iso} (with FEF $> \frac{1}{d}$) can achieve the equality in Eq. (7), then Eq. (11) implies ρ_{iso} is steerable under projective measurement if and only if $W_{\text{Er}}(\rho_{\text{iso}}) \leqslant k_B T \ln \frac{H_d + H_d d - d}{d}$. Being saturated with a resolution of $\delta_\epsilon k_B T \ln 2$, we may interpret $k_B T \ln \frac{H_d + H_d d - d}{d}$ as an *approximate* energy threshold for isotropic states with a possible error of $\delta_\epsilon k_B T \ln 2$, which will be small when the system is large enough. Similar argument applies to other thresholds found in Sec. III B.

B. Theorem 2 and arbitrary pure states

Now let us consider the arbitrary *pure* states $|\phi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ (we adapt the notation $\phi := |\phi\rangle\langle\phi|$). From Ref. [27], we learn that every state $\rho \in \mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$ can be turned into an isotropic state via *quantum twirling*:

$$T(\rho) := \int_{U(d)} (U \otimes U^*)\rho(U \otimes U^*)^\dagger dU, \qquad (17)$$

where dU is the Haar measure representing uniform distribution over the unitary group $U(d)$. Now, the minimal work cost of the quantum twirling on ϕ is given by $k_B T \ln \|T(\Pi_\phi)\|_\infty$ [10], where Π_ϕ is the projector onto the support of the state ϕ. Because ϕ is pure, we have $\Pi_\phi = \phi$. Since $\|\rho_{\text{iso}}\|_\infty = \mathcal{F}(\rho_{\text{iso}})$ and $T(\phi)$ is an isotropic state, the minimal work cost of the quantum twirling is $k_B T \ln \mathcal{F}[T(\phi)]$. By assuming the existence of $\mathcal{P}_{T(|\phi\rangle\langle\phi|)}^{\frac{1}{2}-\epsilon}$, which can extract $W_{\text{Total}}[T(|\phi\rangle\langle\phi|)]$, we use the result for the isotropic state to conclude the following approximation for pure states achieving $\mathcal{F}[T(\phi)] > \frac{1}{d}$:

$$\begin{aligned} W_{\text{Total}}(\phi) &\approx -k_B T \ln \mathcal{F}[T(\phi)] + k_B T \ln \mathcal{F}[T(\phi)]d^2 \\ &= k_B T \ln d^2, \end{aligned} \qquad (18)$$

up to the precision of $\delta_\epsilon k_B T \ln 2$. This is the value predicted by Landauer's principle [5,11,16] (note that, however, this is still an *approximation* because the result for isotropic states is not exact). See Fig. 2 for the schematic interpretation.

V. CONCLUSION

In this work, we try to connect quantum thermodynamics and quantum nonlocality. Consider a given state $\rho \in \mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$ with dimension $d = 2^l$ with $l \in \mathbb{N}$, we prove an inequality which shows that ρ can induce work gain under the local erasure process deterministic in the iid limit if $\mathcal{F}(\rho)$, its fully entangled fraction (FEF), is larger than $\frac{1}{d}$, thereby connecting

work gain under the erasure process to the usefulness of quantum teleportation [27].

By considering work extraction under temperature T as a process mapping the initial state ρ with a fully degenerate Hamiltonian to the Gibbs state in T with the same Hamiltonian (therefore $\frac{\mathbb{I}}{d^2}$), we derive an approximation for the optimal deterministic extractable work in the iid limit, with three prerequisites and an error in the energy scale, which is small in the large system limit. The prerequisites of this approximation also serve as a sufficient condition of the approximate saturation of our first main result. Moreover, a simple picture of the optimal work extraction process deterministic in the iid limit is proposed by this approximation. When it is applicable to the isotropic state, we further obtain a possible thermodynamic interpretation of the FEF: up to an error in the energy scale, the FEF of isotropic states is conceptually equivalent to the minimal work cost (deterministic in the iid limit) of the local erasure process. When we focus on pure states, an improved version of the approximation can be derived. The results we obtained can be a starting point of future research on the interface of quantum thermodynamics and quantum nonlocality.

APPENDIX A: PROOF OF LEMMA 1

Proof. From Ref. [36], we learn that

$$\begin{aligned} S(A|B)_\rho &= \lim_{\epsilon \to 0} \lim_{k \to \infty} \frac{1}{k} H_{\min}^\epsilon(A^{\otimes k}|B^{\otimes k})_{\rho^{\otimes k}} \\ &:= \lim_{\epsilon \to 0} \lim_{k \to \infty} \frac{1}{k} \sup_{\|\eta - \rho^{\otimes k}\|_B < \epsilon} H_{\min}(A^{\otimes k}|B^{\otimes k})_\eta, \quad (A1) \end{aligned}$$

where $\|\rho - \sigma\|_B := \sqrt{2 - 2F(\rho, \sigma)}$ is the *Bures distance* [36] (F is the fidelity defined in Ref. [9]). According to Theorem 2 in Ref. [36], we have

$$H_{\min}(A|B)_\rho = -\log_2[Q(A|B)_\rho d], \qquad (A2)$$

with

$$Q(A|B)_\rho := \max_{\mathcal{E}} \langle \Psi_d^+|(\mathbb{I}_A \otimes \mathcal{E})(\rho)|\Psi_d^+\rangle, \qquad (A3)$$

where $\mathcal{E}: \mathcal{L}(\mathbb{C}^d) \to \mathcal{L}(\mathbb{C}^d)$ is a CPTPM. Substituting the above form, using Lemma 3, and noting the fact $Q(A^{\otimes k}|B^{\otimes k})_{\rho^{\otimes k}} \geqslant \mathcal{F}(\rho^{\otimes k}) \geqslant \mathcal{F}(\rho)^k$, we conclude

$$\begin{aligned} S(A|B)_\rho &= -\lim_{\epsilon \to 0} \lim_{k \to \infty} \frac{1}{k} \log_2[Q(A^{\otimes k}|B^{\otimes k})_{\rho^{\otimes k}} d^k] \\ &\leqslant -\lim_{\epsilon \to 0} \lim_{k \to \infty} \frac{1}{k} \log_2 [\mathcal{F}(\rho)d]^k = -\log_2 \mathcal{F}(\rho)d. \end{aligned}$$

$$(A4)$$

∎

APPENDIX B: NOTES ON THE Q FUNCTION

To start with, let us define the *Hilbert-Schmidt norm* on an operator X as the norm induced by the Hilbert-Schmidt inner product [9]:

$$\|X\|_2 := \sqrt{\mathrm{tr}(X^\dagger X)}. \qquad (\text{B1})$$

First, we prove the following lemma.

Lemma 2. $Q(A|B)_\rho$ is continuous on $\rho \in \mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$ with the Hilbert-Schmidt norm.

Proof. Given $\|\rho - \sigma\|_2 < \epsilon$, there exists a CPTPM $\mathcal{E}_\epsilon$ achieving [we choose $Q(A|B)_\rho = \langle \Psi_d^+|(\mathbb{I}_A \otimes \mathcal{E}_\epsilon)(\rho)|\Psi_d^+\rangle$ when $Q(A|B)_\rho \geqslant Q(A|B)_\sigma$ and $Q(A|B)_\sigma = \langle \Psi_d^+|(\mathbb{I}_A \otimes \mathcal{E}_\epsilon)(\sigma)|\Psi_d^+\rangle$ when $Q(A|B)_\rho \leqslant Q(A|B)_\sigma$]

$$
\begin{aligned}
&|Q(A|B)_\rho - Q(A|B)_\sigma| \\
&\leqslant |\langle \Psi_d^+|(\mathbb{I}_A \otimes \mathcal{E}_\epsilon)(\rho)|\Psi_d^+\rangle - \langle \Psi_d^+|(\mathbb{I}_A \otimes \mathcal{E}_\epsilon)(\sigma)|\Psi_d^+\rangle| \\
&= |\langle \Psi_d^+|(\mathbb{I}_A \otimes \mathcal{E}_\epsilon)(\rho - \sigma)|\Psi_d^+\rangle| \\
&= |\mathrm{tr}[(\mathbb{I}_A \otimes \mathcal{E}_\epsilon^\dagger)(|\Psi_d^+\rangle\langle\Psi_d^+|)(\rho - \sigma)]| \\
&= \frac{1}{d}|\mathrm{tr}[\mathcal{J}(\mathcal{E}_\epsilon^\dagger)(\rho - \sigma)]|, \qquad (\text{B2})
\end{aligned}
$$

where $\mathcal{J}$ is the Choi-Jamiołkowski isomorphism [37]. Now, we use the following generalized Cauchy-Schwarz inequality of operators proved by Bhatia, which holds for arbitrary operators X and Y and arbitrary unitarily invariant norms $\|\cdot\|$ [38,39]:

$$\| |X^\dagger Y|^{\frac{1}{2}} \|^2 \leqslant \|X\| \, \|Y\|, \qquad (\text{B3})$$

where a norm $\|\cdot\|$ is *unitarily invariant* if $\|UXV\| = \|X\|$ holds for all operators X and unitary operators U and V. Since $\|\cdot\|_2$ is unitarily invariant, direct computation shows that

$$
\begin{aligned}
\|X\|_2\|Y\|_2 &\geqslant \| |X^\dagger Y|^{\frac{1}{2}} \|_2^2 \\
&= \mathrm{tr}\big[|X^\dagger Y|^{\frac{1}{2},\dagger}|X^\dagger Y|^{\frac{1}{2}}\big] = \mathrm{tr}|X^\dagger Y|. \qquad (\text{B4})
\end{aligned}
$$

This implies (note that the inequality $|\mathrm{tr}(A)| \leqslant \mathrm{tr}|A|$ holds for *any* operator A[1])

$$
\begin{aligned}
\frac{1}{d}|\mathrm{tr}[\mathcal{J}(\mathcal{E}_\epsilon^\dagger)(\rho - \sigma)]| &\leqslant \frac{1}{d}\mathrm{tr}|\mathcal{J}(\mathcal{E}_\epsilon^\dagger)(\rho - \sigma)| \\
&\leqslant \frac{1}{d}\|\mathcal{J}(\mathcal{E}_\epsilon^\dagger)\|_2\|\rho - \sigma\|_2 \leqslant \frac{1}{d}\|\rho - \sigma\|_2 < \frac{\epsilon}{d}. \qquad (\text{B5})
\end{aligned}
$$

Note that $\|\mathcal{J}(\mathcal{E}_\epsilon^\dagger)\|_2 \leqslant 1$ since $\mathcal{J}(\mathcal{E}_\epsilon^\dagger)$ is a normalized state due to the Choi-Jamiołkowski isomorphism theorem [37]. This completes the proof. ∎

Now, we are in position to prove the following result.

Lemma 3. If $\mathcal{F}(\rho) > \frac{1}{d}$, then for all $k \in \mathbb{N}$ and $\epsilon \in (0,1)$, there exists $0 \leqslant o(\epsilon) \leqslant \log_2(\frac{2\epsilon}{1-2\epsilon} + 1)$ such that

$$\inf_{\|\eta - \rho^{\otimes k}\|_B < \epsilon} \log_2 Q(A^{\otimes k}|B^{\otimes k})_\eta = \log_2 Q(A^{\otimes k}|B^{\otimes k})_{\rho^{\otimes k}} + o(\epsilon). \qquad (\text{B6})$$

Proof. Consider a fixed k value and a fixed $\epsilon \in (0,1)$. We first note the relation $\|\rho - \sigma\|_2 \leqslant 2\|\rho - \sigma\|_B$.[2] This means $\|\eta - \rho^{\otimes k}\|_2 < 2\epsilon$ if $\|\eta - \rho^{\otimes k}\|_B < \epsilon$. When $\|\eta - \rho^{\otimes k}\|_B < \epsilon$, Lemma 2 implies

$$|Q(A^{\otimes k}|B^{\otimes k})_{\rho^{\otimes k}} - Q(A^{\otimes k}|B^{\otimes k})_\eta| \leqslant \frac{1}{d^k}\|\rho^{\otimes k} - \eta\|_2 < \frac{2\epsilon}{d^k}. \qquad (\text{B7})$$

For ρ with $\mathcal{F}(\rho) > \frac{1}{d}$, the function $Q(A^{\otimes k}|B^{\otimes k})_{\rho^{\otimes k}}$ only takes values from the interval $(\frac{1}{d^k}, 1]$. Hence, if $|x - y| < \frac{2\epsilon}{d^k}$ for some $x \in (\frac{1}{d^k}, 1]$ and $y \in (\frac{1-2\epsilon}{d^k}, 1]$ (assume $x \geqslant y$ without loss of generality), we have $|\log_2 x - \log_2 y| < \log_2(\frac{2\epsilon}{d^k y} + 1) < \log_2(\frac{2\epsilon}{1-2\epsilon} + 1)$. Substituting $x = Q(A^{\otimes k}|B^{\otimes k})_{\rho^{\otimes k}}$ and $y = Q(A^{\otimes k}|B^{\otimes k})_\eta$ and considering $\|\eta - \rho^{\otimes k}\|_B < \epsilon$, we conclude

$$\inf_{\|\eta - \rho^{\otimes k}\|_B < \epsilon} \log_2 Q(A^{\otimes k}|B^{\otimes k})_\eta$$
$$= \log_2 Q(A^{\otimes k}|B^{\otimes k})_{\rho^{\otimes k}} - o(\epsilon), \qquad (\text{B8})$$

with the function $o(\epsilon)$ satisfying $0 \leqslant o(\epsilon) \leqslant \log_2(\frac{2\epsilon}{1-2\epsilon} + 1)$. ∎

APPENDIX C: PROOF OF EQ. (13)

Proof. First, we consider a fixed $\epsilon \in (0,1)$. For a given k, we choose $\Delta = \sqrt{k}$ for $\rho^{\otimes k}$ in Supplementary Corollary I.2 in Ref. [11], where their S amounts to our system **S** and Q is now a trivial system (with dimension 1) because we want to study work extraction on **S**. Then this corollary implies the existence of a work extraction process $\mathcal{P}_{\rho^{\otimes k}}^W$ such that $\frac{1}{k}W(\mathcal{P}_{\rho^{\otimes k}}^W) \geqslant k_B T \ln d^2 - [\frac{1}{k}H_{\max}^\epsilon(\rho^{\otimes k}) + \frac{1}{\sqrt{k}}]k_B T \ln 2$, except with a probability of at most $\sqrt{2^{\frac{-\sqrt{k}}{2}} + 12\epsilon}$. Here we use the notation $H_{\max}^\epsilon(\eta) = H_{\max}^\epsilon(S|Q)_\eta$ for the ϵ-smooth max-entropy of S conditional on Q [11]. Using the fact $\lim_{k\to\infty}\frac{1}{k}H_{\max}^\epsilon(\rho^{\otimes k}) = S(\rho) \, \forall \, \epsilon \in (0,1)$, we conclude $\lim_{k\to\infty}P\{\frac{1}{k}W(\mathcal{P}_{\rho^{\otimes k}}^W) \geqslant k_B T \ln d^2 - S(\rho)k_B T \ln 2 - \delta\} = 1 - \sqrt{12\epsilon}$ for all small enough values $\delta > 0$. In other words, this means for every $\delta > 0$, there exists k_δ such that $P\{\frac{1}{k}W(\mathcal{P}_{\rho^{\otimes k}}^W) \geqslant k_B T \ln d^2 - S(\rho)k_B T \ln 2 - \delta\} > 1 - \sqrt{12\epsilon} - \delta \quad \forall \, k \geqslant k_\delta$. For a fixed δ,

---

[1]To prove $|\mathrm{tr}(A)| \leqslant \mathrm{tr}|A|$ for an arbitrary operator A, we first choose the polar decomposition [9] as $A = UJ$, where U is unitary and J is positive. Then one can see it remains to prove $|\mathrm{tr}(UJ)| \leqslant \mathrm{tr}(J)$, where J is nothing but $|A|$. To show this, let us write the spectrum decomposition [9] as $J = \sum_n a_n|\phi_n\rangle\langle\phi_n|$, where $\{|\phi_n\rangle\}$ is an orthonormal basis of the given state space and $a_n \geqslant 0 \, \forall n$. Then direct computation shows $|\mathrm{tr}(UJ)| = |\sum_{n,m}\langle\phi_m|U b_n|\phi_n\rangle\langle\phi_n|\phi_m\rangle| = |\sum_n\langle\phi_n|U|\phi_n\rangle b_n| \leqslant \sum_n |\langle\phi_n|U|\phi_n\rangle|b_n \leqslant \sum_n b_n = \mathrm{tr}(J)$.

[2]Note that $\|\rho - \sigma\|_2 \leqslant \|\rho - \sigma\|_1 \leqslant 2\sqrt{1 - F(\rho,\sigma)^2} \leqslant 2\sqrt{2[1 - F(\rho,\sigma)]} = 2\|\rho - \sigma\|_B$ [9], where $\|A\|_1 := \mathrm{tr}|A|$ is the trace norm. To see the fact $\|\rho - \sigma\|_2 \leqslant \|\rho - \sigma\|_1$, we first apply polar decomposition [9] to write $\rho - \sigma = UJ$, where U is a unitary operator and J is a positive operator. Then we have $\|\rho - \sigma\|_2 = \|J\|_2$ and $\|\rho - \sigma\|_1 = \|J\|_1$. Choose the spectrum decomposition [9] as $J = \sum_n a_n|\phi_n\rangle\langle\phi_n|$, where $\{|\phi_n\rangle\}$ is an orthonormal basis and $a_n \geqslant 0 \, \forall n$. One can verify that $\|J\|_2^2 = \sum_n a_n^2 \leqslant (\sum_n a_n)^2 = \|J\|_1^2$, which proves the desired result.

we let $\epsilon \to 0$ and obtain $P\{\frac{1}{k}W(\mathcal{P}_{\rho^{\otimes k}}^{\mathrm{W}}) \geqslant k_B T \ln d^2 - S(\rho)k_B T \ln 2 - \delta\} > 1 - \delta \quad \forall \ k \geqslant k_\delta$. This implies $k_B T \ln d^2 - S(\rho)k_B T \ln 2 - \delta \leqslant W_{\mathrm{Total}}(\rho)$ for all $\delta > 0$. Letting $\delta \to 0$ will result in the desired inequality. ∎

APPENDIX D: PROOF OF THEOREM 2

We first prove the following proposition, which is a general version of Theorem 2 [recall we define $S_{\min}(\rho) := \min\{S(\rho_A), S(\rho_B)\}$].

Proposition 1. Given $0 < \epsilon \leqslant \frac{1}{2}$ and $\delta > \delta_\epsilon := -3\ln\epsilon$. If (i) $\mathcal{F}(\rho) > \frac{1}{d}$, (ii) there exists $\mathcal{P}_\rho^{\frac{1}{2}-\epsilon}$ which can extract $W_{\mathrm{Total}}(\rho)$, and (iii) $-\delta_\epsilon < \log_2 \|\rho\|_\infty - [\log_2 \mathcal{F}(\rho)d - S_{\min}(\rho)] < \delta - \delta_\epsilon$, then we have

$$\left| \frac{W_{\mathrm{Total}}(\rho)}{k_B T \ln 2} - [\log_2 \mathcal{F}(\rho)d^3 - S_{\min}(\rho)] \right| < \delta. \quad \text{(D1)}$$

Proof. From Lemma 1, Eq. (13), and Refs. [12,16], we obtain

$$k_B T \ln d^2 - [H_{\min}^\epsilon(\rho) + 3\ln\epsilon]k_B T \ln 2$$
$$> W_{\mathrm{Total}}(\rho) \geqslant k_B T \ln d^2 - S(\rho)k_B T \ln 2$$
$$\geqslant k_B T \ln \mathcal{F}(\rho)d^3 - S_{\min}(\rho)k_B T \ln 2, \quad \text{(D2)}$$

where [12]

$$H_{\min}^\epsilon(\rho) := \sup_{\|\eta - \rho\|_B < \epsilon} (-\log_2 \|\eta\|_\infty) \geqslant -\log_2 \|\rho\|_\infty. \quad \text{(D3)}$$

This implies

$$\log_2 d^2 + \log_2 \|\rho\|_\infty - 3\ln\epsilon > \frac{W_{\mathrm{Total}}(\rho)}{k_B T \ln 2}$$
$$\geqslant \log_2 \mathcal{F}(\rho)d^3 - S_{\min}(\rho). \quad \text{(D4)}$$

Hence, a sufficient condition for the inequality $\left|\frac{W_{\mathrm{Total}}(\rho)}{k_B T \ln 2} - [\log_2 \mathcal{F}(\rho)d^3 - S_{\min}(\rho)]\right| < \delta$ reads

$$0 < \log_2 \|\rho\|_\infty - [\log_2 \mathcal{F}(\rho)d - S_{\min}(\rho)] + \delta_\epsilon < \delta, \quad \text{(D5)}$$

where $\delta_\epsilon := -3\ln\epsilon > 0$ is the best precision that we may have, since it is the lower bound of all the allowed δ values. Note that being larger than 0 is necessary to guarantee no contradiction. ∎

We are now in position to prove Theorem 2.

Proof. To begin with, note that $W_{\mathrm{Er}}(\rho) \geqslant k_B T \ln \mathcal{F}(\rho)d$ for state ρ with $\mathcal{F}(\rho) > \frac{1}{d}$ from Theorem 1, and $W_{\mathrm{Total}}(\rho) \leqslant \delta k_B T \ln 2 + k_B T \ln 2[\log_2 \mathcal{F}(\rho)d^3 - S_{\min}(\rho)]$ if it is possible to choose $\mathcal{P}_\rho^{\frac{1}{2}-\epsilon}$ to extract $W_{\mathrm{Total}}(\rho)$ and $\log_2 \|\rho\|_\infty = \log_2 \mathcal{F}(\rho)d - S_{\min}(\rho)$ (hence we can choose $\delta \approx \delta_\epsilon := -3\ln\epsilon$), which is due to Theorem 2. Then direct computation shows [choose $S_{\min}(\rho) = S(\rho_B)$ without loss of generality]

$$0 \leqslant W_{\mathrm{Total}}(\rho) - [k_B T \ln d^2 - S_{\min}(\rho)k_B T \ln 2 + W_{\mathrm{Er}}(\rho)]$$
$$\leqslant \delta k_B T \ln 2 + k_B T \ln \mathcal{F}(\rho)d - W_{\mathrm{Er}}(\rho)$$
$$\leqslant \delta k_B T \ln 2 \approx \delta_\epsilon k_B T \ln 2, \quad \text{(D6)}$$

where the first inequality follows from the fact that $k_B T \ln d^2 - S_{\min}(\rho)k_B T \ln 2 + W_{\mathrm{Er}}(\rho)$ can be interpreted as the deterministic work gain (in the iid limit) of a particular work extraction process (see Fig. 1), thereby being smaller than the optimal one, i.e., $W_{\mathrm{Total}}(\rho)$. This implies, up to the energy scale $\delta_\epsilon k_B T \ln 2$, the desired results. ∎

[1] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).

[2] J. S. Bell, Physics **1**, 195 (1964).

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

[4] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961).

[5] R. Alicki, M. Horodecki, P. Horodecki, and R. Horodecki, Open Syst. Inf. Dyn. **11**, 205 (2004).

[6] M. Horodecki and J. Oppenheim, Nat. Commun. **4**, 2059 (2013).

[7] P. Skrzypczyk, A. J. Short, and S. Popescu, Nat. Commun. **5**, 4185 (2014).

[8] J. Åberg, Nat. Commun. **4**, 1925 (2013).

[9] M. A. Nielson and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th anniv. ed. (Cambridge University Press, Cambridge, England, 2010).

[10] P. Faist, F. Dupuis, J. Oppenheim, and R. Renner, Nat. Commun. **6**, 7669 (2015).

[11] L. d. Rio, J. Åberg, R. Renner, O. Dahlsten, and V. Vedral, Nature (London) **474**, 61 (2011).

[12] O. C. O. Dahlsten, R. Renner, E. Rieper, and V. Vedral, New J. Phys. **13**, 053015 (2011).

[13] K. V. Hovhannisyan, M. Perarnau-Llobet, M. Huber, and A. Acín, Phys. Rev. Lett. **111**, 240401 (2013).

[14] M. Perarnau-Llobet, K. V. Hovhannisyan, M. Huber, P. Skrzypczyk, N. Brunner, and A. Acín, Phys. Rev. X **5**, 041011 (2015).

[15] K. Funo, Y. Watanabe, and M. Ueda, Phys. Rev. A **88**, 052319 (2013).

[16] J. Oppenheim, M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **89**, 180402 (2002).

[17] F. N. C. Paraan and A. Silva, Phys. Rev. E **80**, 061130 (2009).

[18] L. Fusco, M. Paternostro, and G. D. Chiara, Phys. Rev. E **94**, 052122 (2016).

[19] K. Korzekwa, M. Lostaglio, J. Oppenheim, and D. Jennings, New J. Phys. **18**, 023045 (2016).

[20] A. Misra, U. Singh, S. Bhattacharya, and A. K. Pati, Phys. Rev. A **93**, 052335 (2016).

[21] G. L. Giorgy and S. Campbell, J. Phys. B: At. Mol. Opt. Phys. **48**, 035501 (2015).

[22] N. Brunner, D. Cavalcanti, S. Pironio, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).

[23] E. Schrödinger, Proc. Cambridge Philos. Soc. **31**, 555 (1935); **32**, 446 (1936).

[24] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Phys. Rev. Lett. **98**, 140402 (2007).

[25] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).

[26] S. Popescu, Phys. Rev. Lett. **72**, 797 (1994).

[27] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **60**, 1888 (1999).

[28] S. Albeverio, S.-M. Fei, and W.-L. Yang, Phys. Rev. A **66**, 012301 (2002).

[29] M.-J. Zhao, Z.-G. Li, S.-M. Fei, and Z.-X. Wang, J. Phys. A: Math. Theor. **43**, 275203 (2010).

[30] D. Cavalcanti, A. Acín, N. Brunner, and T. Vértesi, Phys. Rev. A **87**, 042104 (2013).

[31] C.-Y. Hsieh, Y.-C. Liang, and R.-K. Lee, Phys. Rev. A **94**, 062120 (2016).

[32] M. T. Quintino, N. Brunner, and M. Huber, Phys. Rev. A **94**, 062123 (2016).

[33] R. Renner, Security of Quantum Key Distribution, Ph.D. dissertation, ETH Zürich, Switzerland (2005).

[34] M. L. Almeida, S. Pironio, J. Barrett, G. Tóth, and A. Acín, Phys. Rev. Lett. **99**, 040403 (2007).

[35] R. Renner and S. Wolf, in *Proceedings of International Symposium on Information Theory, ISIT 2004* (IEEE, New York, 2004), p. 232.

[36] R. König, R. Renner, and C. Schaffner, IEEE Trans. Inf. Theory **55**, 4337 (2009).

[37] A. Jamiołkowski, Rep. Math. Phys. **3**, 275 (1972).

[38] R. Bhatia, J. Oper. Theory **19**, 129 (1988).

[39] R. Bhatia and C. Davis, Linear Algebra Appl. **223-224**, 119 (1995).