Coherence distillation machines are impossible in quantum thermodynamics

Iman Marvian¹

¹Departments of Physics & Electrical and Computer Engineering, Duke University, Durham, North Carolina 27708, USA

The role of coherence in quantum thermodynamics has been extensively studied in the recent years and it is now well-understood that coherence between different energy eigenstates is a resource independent of other thermodynamics resources, such as work. A fundamental remaining open question is whether the laws of quantum mechanics and thermodynamics allow the existence a coherence distillation machine, i.e. a machine that, by possibly consuming work, obtains pure coherent states from mixed states, at a nonzero rate. This question is related to another fundamental question: Starting from many copies of noisy quantum clocks which are (approximately) synchronized with a reference clock, can we distill synchronized clocks in pure states, at a non-zero rate? In this paper we study quantities called coherence cost and distillable coherence, which determine the rate of conversion of coherence in a standard pure state to general mixed states, and vice versa, in the context of quantum thermodynamics. We find that the coherence cost of any state (pure or mixed) is determined by its Quantum Fisher Information (QFI), thereby revealing a novel operational interpretation of this central quantity of quantum metrology. On the other hand, we show that, surprisingly, distillable coherence is zero for typical (full-rank) mixed states. Hence, we establish the impossibility of coherence distillation machines in quantum thermodynamics, which can be compared with the impossibility of perpetual motion machines or cloning machines. To establish this result, we introduce a new additive quantifier of coherence, called the *purity* of coherence, and argue that its relation with QFI is analogous to the relation between the free and total energies in thermodynamics.

What are the fundamental limits of nature on manipulation of quantum clocks? Suppose we have multiple clocks, all synchronized with the same reference clock, which are affected by noise. Then, by averaging the time read from these clocks we can obtain a more accurate estimate of the current time according to the reference clock. In other words, we can *distill* a less noisy clock from several noisy clocks. What are the limits of this distillation process for quantum clocks? Can we distill quantum clocks in *pure* states from those in mixed states, at a nonzero rate?

Interestingly, this question is related to another fundamental question about the manipulation of quantum coherence in quantum thermodynamics. It is now well-understood that coherence between different energy eigenstates is a resource in quantum thermodynamics, independent of other resources such as work, and can be used to implement operations which are otherwise impossible [1–4]. A fundamental open question in this context is whether the laws of quantum mechanics and thermodynamics allow the existence a coherence distillation machine, i.e. a machine that consumes work to obtain pure coherent states from mixed ones at a nonzero rate (See Fig.1). The connection between these two questions arises from the fact that the minimum requirement for a system to be a clock is to be in a state which contains coherence with respect to the energy-eigenbasis; otherwise, the system will be time-independent, and hence useless as a clock.

In this paper we investigate the coherence *distillation* and *formation* processes in the context of quantum thermodynamics. The latter process, which can be thought as the time reversal of the former, is the process by which one prepares an arbitrary mixed coherent state by consuming pure coherent states. We study quantities called the *distillable coherence* and the *coherence cost*, which determine the rate of conversion of a general mixed coherent state to a standard pure state, and vice versa. These quantities can be thought as the counterparts

of the distillable entanglement and the entanglement cost in the resource theory of entanglement [5], and finding them has been an open question [6, 7] in the resource theory of quantum thermodynamics (athermality) and the closely related resource theory of asymmetry. In this paper, we first show that the coherence cost of preparing a quantum state is determined by its Quantum Fisher Information (QFI) [8–10]. This reveals a novel operational interpretation of this central quantity of quantum metrology [8–12]. On the other hand, while for pure states the distillable coherence is also determined by the QFI, surprisingly, we find that for a typical (full-rank) mixed state this quantity is zero. Using the terminology of quantum resource theories [5, 13–15], this means that a typical state is a bound resource, meaning that its creation requires consuming the resource, in this case coherence, and yet we cannot distill any pure coherence from it. We conclude that, the hypothetical coherence distillation machine depicted in Fig.1, which distills coherent pure states from a general mixed state, is forbidden by the laws of nature, in the same vein that perpetual motion machine or cloning machine [16, 17] are forbidden.

I. COHERENCE AND QUANTUM CLOCKS

A quantum clock is characterized by its state and Hamiltonian, which usually generates a periodic time evolution [18–25]. By definition, state of a clock should be time-dependent. Therefore, when we say a clock with Hamiltonian H is in state ρ , we actually mean its state is ρ at a particular time, say t=0, with respect to a reference clock. Then, at an arbitrary time t the state of clock is $e^{-iHt}\rho e^{iHt}$ (Throughout this paper we assume $\hbar=1$). Here, we focus on the systems with periodic dynamics, whose period is equal to a fixed (but arbitrary) parameter τ , such that $\tau=\inf_t\{t>0:e^{-iHt}\rho e^{iHt}=\rho\}$. Otherwise the state and Hamiltonian are completely arbi-

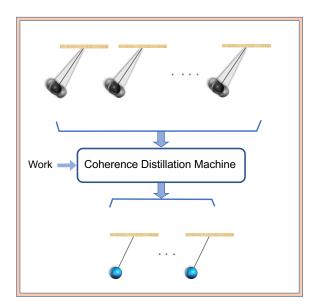


FIG. 1: A hypothetical "Coherence Distillation Machine" for distilling coherence with respect to the energy eigenbasis: It consumes work and obtains pure coherent states from mixed states at a non-zero rate, or equivalently, purifies quantum clocks. The clock could be, for instance, a harmonic oscillator (pendulum), or a two-level system. Is this hypothetical machine consistent with the laws of quantum mechanics and thermodynamics?

trary. Examples include a two-level system with Hamiltonian $H=\pi\sigma_z/\tau$ or a Harmonic oscillator with frequency τ^{-1} , in any pure state except the energy eigenstates.

A quantum clock can have a wide variety of functionalities. For instance, it can be used for reading the current time, or as a time reference for preparing other quantum systems in the desired states; when we shine a laser beam to an atom and prepare it in the superposition of two energy levels, the laser beam plays the role of a clock. Due to this variety, it turns out that there is no single figure of merit characterizing the overall usefulness of a quantum clock or quantifying its information content about time. Rather, a more appropriate approach for characterizing clocks is to use a *resource theoretic* method [13–15, 26, 27], in which quantum clocks are classified based on how they can be manipulated and transformed to each other.

To study the manipulation of clocks in a systematic way, we first identify the set of transformations which can be implemented on quantum systems without knowing the current time, that is without having access to a synchronized clock. In the language of quantum resource theories, these are the *free* operations. An example of this type of transformations is energy-conserving unitaries, i.e. any unitary U which commutes with the intrinsic Hamiltonian of clock. On the other hand, there are also non-energy conserving transformations which can be implemented without interacting with a reference clock. For instance, preparing a quantum system in any state ρ which is invariant under time translations, such that $\rho = e^{-iHt}\rho e^{iHt}$ for all times t, does not require any interac-

tion with a synchronized clock.

In general, it turns out that the lack of access to a synchronized clock restricts the set of possible state transformations to those which can be implemented by time-Translationally Invariant (TI) operations, i.e. those which satisfy the *covariance condition*,

$$e^{-iH_{\text{out}}t} \mathcal{E}_{\text{TI}}(\sigma) e^{iH_{\text{out}}t} = \mathcal{E}_{\text{TI}}(e^{-iH_{\text{in}}t}\sigma e^{iH_{\text{in}}t}),$$
 (1)

for all density operators σ and all times t [20, 28–30]. Here, \mathcal{E}_{TI} is a linear transformation mapping density operators of the input to the output, which in general can be different systems, and H_{in} and H_{out} are their corresponding Hamiltonians (Note that manipulating a clock may involve a change in its state, Hamiltonian or both). As a physically realizable transformation, \mathcal{E}_{TI} should map density operators to density operators, and more specifically should be trace-preserving and completely-positive [31, 32]. TI operations are precisely those operations which can be implemented without having access to (or interacting with) a reference clock, and therefore they define a natural framework to study the manipulation of clocks [20, 28–30]. Note that the time-translation symmetry in Eq.(1) guarantees that TI operations can be implemented on the system without interfering with its intrinsic time evolution generated by its own Hamiltonian.

In this paper we consider transformations on non-interacting composite systems. Two closed systems A and B are not interacting with each other if, and only if, their total Hamiltonian is the sum of the Hamiltonians of the individual systems, i.e. $H_A \otimes I_B + I_A \otimes H_B$, where $I_{A/B}$ and $H_{A/B}$ are, respectively, the identity operators and the Hamiltonians of A and B. We also consider n copies of a system with state ρ and Hamiltonian H, by which we mean a composite system with state $\rho^{\otimes n}$ and the total Hamiltonian $\sum_i H^{(i)}$, where $H^{(i)} = I^{\otimes (i-1)} \otimes H \otimes I^{\otimes (n-i-1)}$. Note that, in general, to implement a TI operation on non-interacting systems we may need to turn on an interaction between them and let them exchange energy.

Clearly, under the restriction to TI operations some state conversions are impossible. In particular, starting from state ρ which is invariant under time translations one cannot obtain a state which breaks this symmetry. Note that a state breaks time-translation symmetry if, and only if, it contains coherence, i.e. off-diagonal terms with respect to the energy eigenbasis. In this sense, coherence becomes a resource which cannot be generated under TI operations. Another example of state conversions which are forbidden by time-translation symmetry is transforming an input system to an output with a higher period. In particular, a system with period τ can be only transformed to systems with period τ/k , for a positive integer k>0.

In general, finding the consequences of this type of restriction on state transformations is the subject of the resource theory of asymmetry, which studies asymmetry (or symmetry-breaking) with respect to a general symmetry and arises naturally from the study of quantum clocks and reference frames (See e.g. [20, 28, 29, 33–52]). Interestingly, as we will see in the following, this resource theory also provides a natu-

ral framework for understanding manipulation of coherence in quantum thermodynamics (See [7] and [53, 54] for further discussions on different resource theories of coherence, and the relation between asymmetry and coherence).

A. Coherence and Quantum Thermodynamics

Thermodynamic resource theories study thermodynamic concepts in a resource-theoretic framework, similar to the entanglement and asymmetry resource theories, and are often defined based on the principle that any non-thermal state is a resource [14, 26, 55–60]. A prominent version of these resource theories is defined based on the notion of *thermal operations* [55, 57], that is the set of operations which can be implemented on a given quantum system if one only has access to thermal baths in a fixed temperature β^{-1} , and can only implement unitary transformations which conserve the total energy, such that

$$\Lambda_{\text{thm}}(\rho) = \text{Tr}_B U(\rho \otimes \gamma_\beta) U^{\dagger}, \tag{2}$$

where B denotes the thermal bath which is initially in the thermal state $\gamma_{\beta} = e^{-\beta H_B}/\mathrm{Tr}(e^{-\beta H_B})$, and unitary U is energy-conserving, in the sense that $[U, I_S \otimes H_B + H_S \otimes I_B] = 0$, where H_S and H_B are, respectively the input system and the bath Hamiltonians. Note that, in general, the input and output systems can be different. Under thermal operations, an input system which is initially in a thermal state can only be mapped to a thermal state, as it is expected from the second law of thermodynamics [1, 2].

Any thermal operation is invariant under time translations, and therefore is a TI operation satisfying Eq.(1) [1, 2]. However, not all TI operations are a thermal operation; for example, the transformation which maps the input to an energy eigenstate is a TI, but not thermal, operation. On the other hand, interestingly, it turns out that by consuming enough work (free energy) one can implement any TI operation using an energy-conserving unitary, which is a thermal operation: Any TI operation \mathcal{E}_{TI} on a system S with Hamiltonian H_S can be implemented by coupling the system to an auxiliary system, e.g. a battery or work reservoir, with Hamiltonian H_{aux} , such that

$$\mathcal{E}_{\text{TI}}(\sigma) = \text{Tr}_{\text{aux}} U(\sigma \otimes |E_0\rangle \langle E_0|_{\text{aux}}) U^{\dagger} , \qquad (3)$$

where (i) the initial state $|E_0\rangle_{\rm aux}$ of the auxiliary system is an eigenstate of Hamiltonian $H_{\rm aux}$, and (ii) the unitary U that couples it to the system S conserves the total energy $H_{\rm tot}=H_S\otimes I_{\rm aux}+I_S\otimes H_{\rm aux}$, i.e. $[U,H_{\rm tot}]=0$ [61]. This observation further clarifies the fact that coherence and work (free energy) are distinct resources.

Using the notions of TI and thermal operations we can provide a formal framework for the coherence distillation machine in Fig.(1): a coherence distillation machine should be a TI operation, or equivalently, should be a thermal operation supplemented by an arbitrary amount of work, that is an auxiliary system initially in an energy eigenstate (See Eq.(3)). The purpose of the coherence distillation machine is to distill

coherence in the form of pure states, as formally defined in Sec.V A.

In summary, in this section we highlighted two different motivations to study state conversions under TI operations: (i) manipulation of quantum clocks, and (ii) manipulation of coherence in quantum thermodynamics. See [53] for further discussion on applications of TI operations in areas such as quantum metrology [11, 12].

II. REVERSIBLE TRANSFORMATION BETWEEN PURE STATES

Consider many copies of a system, e.g. a quantum clock, with Hamiltonian H_1 and pure state ψ_1 with a time evolution with period τ (Recall that by copies of a system we mean non-interacting systems, each with Hamiltonian H_1 and state ψ_1). Is it possible to transform these systems to many copies of another system with the same period τ , in pure state ψ_2 and Hamiltonian H_2 , using only TI operations? In practice, the exact transformations are often impossible and physically intractable. Therefore, we can allow a small error ϵ , provided that it vanishes in the limit of infinite copies.

This question has been studied before in [28, 39, 40]. Here, in theorem 1, we present a more general and stronger version of the previous result (See the Supplementary Material (SM) for the proofs of this theorem and all other results in the paper). Most importantly, this theorem establishes convergence in trace distance $D(\rho,\sigma)=\|\rho-\sigma\|_1$, which, by the Helstrom's theorem [8, 31, 32], is the relevant notion of convergence from an operational point of view (Such convergence does not follow from the previous argument of [28, 39, 40]). Furthermore, unlike theorem 1, the previous result only establishes reversible conversion between a limited family of pure states (namely states with *gapless* energy spectrum), and also does not apply to systems in infinite-dimensional Hilbert spaces, such as harmonic oscillators in pure Gaussian states, which are of important practical interest.

In the following, $V_H(\psi) = \langle \psi | H^2 | \psi \rangle - \langle \psi | H | \psi \rangle^2$ denotes the energy variances of states.

Theorem 1. (Based on [28, 39, 40]) Suppose the period of a system with Hamiltonian H_1 and pure state ψ_1 is equal to the period of another system with Hamiltonian H_2 and pure state ψ_2 , and for both systems the third moments of energy are finite. Then, using TI operations the asymptotic transformation

$$\psi_1^{\otimes n} \xrightarrow{TI} \stackrel{\epsilon}{\approx} \psi_2^{\otimes \lceil Rn \rceil} \quad \text{as } n \to \infty \;, \; \epsilon \to 0 \;,$$

is possible with vanishing error ϵ , given that rate $R \leq V_{H_1}(\psi_1)/V_{H_2}(\psi_2)$ and is impossible with rate $R > V_{H_1}(\psi_1)/V_{H_2}(\psi_2)$.

Theorem 1 essentially means that in the asymptotic regime all oscillators (with the same frequency) in pure states are equivalent resources, in the sense that by adding or absorbing sufficient amount of energy their coherence content, or equivalently, their information content about time, can be converted from one form to another, in a reversible fashion (Note that the

maximal rate from ψ_1 to ψ_2 is the inverse of the rate from ψ_2 to ψ_1 , and therefore the transformation is reversible). Consequently, in this regime the usefulness of a clock can be quantified by a single number, namely its energy variance. Furthermore, if we know the optimal rate of conversation of a general (possibly mixed) state to a certain pure state, we can also determine the optimal rate of its conversion to any other pure state. Therefore, we can pick a standard clock-bit (coherence-bit) or c-bit with period τ and quantify the amount of resource of a general state relative to this standard. A convenient choice is a two-level system with Hamiltonian $H_{c\text{-bit}} = \pi \sigma_z/\tau$ and state $|\Phi\rangle_{c\text{-bit}} = (|0\rangle + |1\rangle)/\sqrt{2}$, which has the energy variance π^2/τ^2 .

Theorem 1 only applies to pure states. In the rest of this paper we consider two variants of this scenario where the output or input are mixed states (The version with mixed input and pure output corresponds to the coherence distillation machine). But, first we need to understand the physical significance of the energy variance in this theorem.

A. Quantum Fisher Information

Quantum Fisher Information (QFI) is a central quantity of quantum metrology and estimation theory [8–10, 62, 63]. Consider the family of states $\{e^{-iHt}\rho e^{iHt}\}_t$ corresponding to the time-translated versions of ρ . The QFI (for the time parameter t) for this family of states is defined by

$$F_H(\rho) = 2\sum_{j,k} \frac{(p_j - p_k)^2}{p_j + p_k} |\langle \psi_j | H | \psi_k \rangle|^2,$$
 (4)

where $\rho=\sum_j p_j |\psi_j\rangle\langle\psi_j|$ is the spectral decomposition of ρ . Equivalently, QFI can be expressed as the second derivative of the fidelity of states ρ and $e^{-iHt}\rho e^{iHt}$ with respect to parameter t [64]. QFI is a natural generalization of the classical Fisher information, and similar to its classical counterpart satisfies the Cramer-Rao bound [8–10, 62, 63], $\Delta t \times F_H(\rho) \geq 1$, where Δt is the variance of any unbiased estimator of the time parameter t.

QFI has several important properties [8–10, 62, 63]: (i) Faithfulness: It is zero if, and only if, state is incoherent, i.e. diagonal in the energy eigenbasis. (ii) Monotonicity: It is non-increasing under any TI operation \mathcal{E}_{TI} , i.e. $F_H(\mathcal{E}_{TI}(\rho)) \leq F_H(\rho)$. In particular, it remains invariant under energy-conserving unitaries. (iii) Additivity: For a composite non-interacting system with the total Hamiltonian $H_{tot} = H_1 \otimes I_2 + I_1 \otimes H_2$, QFI is additive for uncorrelated states, i.e. $F_{H_{tot}}(\rho_1 \otimes \rho_2) = F_{H_1}(\rho_1) + F_{H_2}(\rho_2)$. (iv) Convexity: For any $0 \leq p \leq 1$ and states ρ and σ , $F_H(p\rho + (1-p)\sigma) \leq pF_H(\rho) + (1-p)F_H(\sigma)$.

QFI is an example of measures of asymmetry (with respect to a translational symmetry) which can be obtained from measures of distinguishability using the approach discussed in [65]. A measure of asymmetry is a function which quantifies the amount of asymmetry (or symmetry-breaking) relative to a symmetry, and by definition, is zero on states which do not break the symmetry and satisfies the monotonicity condi-

tion (ii). This interpretation of QFI as measure of asymmetry [66, 67] has lead to new applications of QFI [66], e.g. in the context of Quantum thermodynamics [68].

For pure states, QFI reduces to the energy variance,

$$F_H(\psi) = 4V_H(\psi) . (5)$$

This suggests a simple interpretation of theorem 1: In the asymptotic regime, a coherent pure state with period τ can be transformed to any other pure state with the same period, with any rate R which is consistent with the monotonicity of QFI (See also [67]). This is consistent with the fact that, roughly speaking, TI operations do not have any information about the time parameter, and therefore, they cannot increase the sensitivity of states with respect to time. This interpretation suggests that to generalize theorem 1 to mixed states, the role of variance should be replaced by QFI. Indeed, we will see that this intuition is partially correct.

III. OPERATIONAL INTERPRETATION OF FISHER INFORMATION

Next, we study the cost of creating a general mixed state from pure coherent states using TI operations, a process known as *formation*. This amounts to a generalization of theorem 1 to the case where the final state ψ_2 is replaced by an arbitrary mixed state with period τ . But, first, we start with the single-shot version of this problem.

A. Single-shot regime

Purification is a deep principle of quantum mechanics stating that any system in a mixed state can be thought as a subsystem of a larger system in a pure state [31, 32, 69]. For pure states, we saw that the energy variance, or equivalently, QFI, provides a quantification of the coherence content of state. Therefore, a natural way to quantify the coherence content of a mixed state ρ is to find the minimum QFI of a pure state which purifies ρ .

Suppose to prepare system S with Hamiltonian H_S in a general mixed state ρ , we prepare S and an *auxiliary* system A, which is not interacting with S, in a pure state $|\Phi_{\rho}\rangle_{SA}$ with partial trace $\mathrm{Tr}_A(|\Phi_{\rho}\rangle\langle\Phi_{\rho}|_{SA})=\rho$. This means that by discarding A we prepare S in state ρ . We are interested in the energy variance, or, equivalently, QFI of the purifying state $|\Phi_{\rho}\rangle_{SA}$ with respect to the total Hamiltonian $H_{\mathrm{tot}}=H_S\otimes I_A+I_S\otimes H_A$, where H_A is the auxiliary system Hamiltonian. Considering all such purifications $|\Phi_{\rho}\rangle_{SA}$ and all Hamiltonians H_A , what is the minimum possible energy variance $V_{H_{\mathrm{tot}}}(\Phi_{\rho})$ (or, equivalently QFI) of the purifying state $|\Phi_{\rho}\rangle_{SA}$ with respect to the total Hamiltonian H_{tot} ?

Answering this question leads us to a new property of QFI.

Theorem 2. QFI of state ρ of system S is four times the minimum energy variance of all purifications of ρ with auxiliary

systems not interacting with S, i.e.

$$F_H(
ho) = \min_{\Phi_
ho, H_A} F_{H_{tot}}(\Phi_
ho) = 4 imes \min_{\Phi_
ho, H_A} V_{H_{tot}}(\Phi_
ho) , \quad (6)$$

where the minimization is over all pure states $|\Phi_{\rho}\rangle_{SA}$ satisfying $Tr_A(|\Phi_{\rho}\rangle\langle\Phi_{\rho}|_{SA}) = \rho$, and all Hamiltonians of A.

Note that, in contrast to most other previously known properties of QFI, such as additivity, monotonicity, and Cramer-Rao bound, this new property of QFI, which is related to the purification principle, does not have any classical counterpart.

A remarkable corollary of this result is that there exists pure bi-partite states for which by discarding a subsystem, the QFI does not decrease at all, even though, as we will see later, the discarding process is irreversible. In other words, for some applications, the composite system SA is a more useful clock than system SA alone, but they have the same QFI.

B. Asymptotic regime: Coherence cost

Consider a system with state ρ and Hamiltonian H with period τ . The coherence cost of this state (with respect to TI operations) is the minimal rate at which c-bits with period τ , i.e. two level systems with state $|\Phi\rangle_{\text{c-bit}} = (|0\rangle + |1\rangle)/\sqrt{2}$ and Hamiltonian $H_{\text{c-bit}} = \pi \sigma_z/\tau$, have to be consumed for preparing ρ .

Definition 3. The coherence cost (w.r.t. TI operations) of a system with state ρ and Hamiltonian H is defined as

$$C_c^{\mathit{TI}}(\rho) = \inf R : \Phi_{c\text{-bit}}^{\otimes \lceil Rn \rceil} \xrightarrow{TI} \stackrel{\epsilon}{\approx} \rho^{\otimes n} \text{ as } n \to \infty, \epsilon \to 0 \ .$$

A possible strategy for generating a state ρ is to prepare $|\Phi_{\rho}\rangle$, the optimal purification of ρ in theorem 2, and then discard the auxiliary system. According to theorem 2, the QFI of this pure state is equal to the QFI of ρ . Then, combining this with theorem 1, one may expect that using this approach $\Phi_{\text{c-bit}}$ can be converted to ρ with the rate $F_{\text{c-bit}}/F_H(\rho)$, indicating that $C_c^{\text{TI}}(\rho) \leq F_H(\rho)/F_{\text{c-bit}}$, where $F_{\text{c-bit}}$ is the QFI of the standard state $\Phi_{\text{c-bit}}$. However, it turns out that, in general, for the optimal pure state $|\Phi_{\rho}\rangle$ and Hamiltonian H_A in theorem 2, which achieve the lowest energy variance, the period of dynamics is not equal to τ , the period of ρ . Therefore, $|\Phi_{\rho}\rangle$ can not be obtained from a standard pure state with period τ (i.e. theorem 1 assumptions are not satisfied).

Therefore, we use a different strategy: state ρ can also be generated by preparing an ensemble of pure states $\{p_i, |\phi_i\rangle\}$ such that $\sum_i p_i |\phi_i\rangle\langle\phi_i| = \rho$. Interestingly, it turns out that there exists an optimal ensemble for which the average QFI is equal to the QFI of state ρ .

Theorem 4. (Yu-Toth-Petz [70, 71]) *QFI is four times the* convex roof *of the variance, i.e.*

$$F_H(\rho) = \min_{\{p_i, \phi_i\}} \sum_i p_i F_H(\phi_i) = 4 \times \min_{\{p_i, \phi_i\}} \sum_i p_i V_H(\phi_i) ,$$
(7)

where the minimization is over the set of all ensembles of pure states $\{p_i, \phi_i\}$ satisfying $\sum_i p_i |\phi_i\rangle \langle \phi_i| = \rho$.

This theorem was originally conjectured (and proven in a special case) by Toth and Petz [70] and was proven later by Yu [71]. In the SM we show that this theorem follows easily from theorem 2 together with the monotonicity of QFI under TI operations, thus providing a new proof of this result. Note that in analogy with the entanglement theory, the right-hand side of Eq.(7) can be called *coherence of formation* (w.r.t. TI operations) [72]. Using this theorem together with an extension of theorem 1, in the SM we show that $C_c^{\rm TI}(\rho) \leq F_H(\rho)/F_{c\text{-bit}}$.

It turns out that this bound holds as equality: In the SM we show that, even at the presence of error ϵ , provided that it vanishes in the limit $n\to\infty$, the rate of any transformation should be bounded by the ratio of the QFI of the input to the output states (Although this might be expected from the monotonicity and additivity of QFI, note that allowing error ϵ can change QFI by order $n^2\epsilon$. In other words, QFI is not asymptotically continuous). This proves the following theorem:

Theorem 5. Consider a system with Hamiltonian H and state ρ in a finite-dimensional Hilbert space, which has periodic dynamics with period τ . The coherence cost (w.r.t. TI operations) of this system is equal to the ratio of Quantum Fisher Information for state ρ and for the standard pure coherent state with period τ , i.e.

$$C_c^{TI}(\rho) = \frac{F_H(\rho)}{F_{c-bit}} = \left(\frac{\tau}{2\pi}\right)^2 \times F_H(\rho) . \tag{8}$$

As we saw before, by consuming work, any TI operation can be implemented by thermal operations. Therefore, from the point of view of quantum thermodynamics, this result means that by consuming work we can transform the standard state $\Phi_{\text{c-bit}}$ to state ρ with the optimal rate $F_H(\rho)/F_{\text{c-bit}}$.

This result together with our single-shot result in Theorem 2 establish a novel operational interpretation of QFI as the coherence cost in a general thermodynamic context.

IV. PURITY OF COHERENCE

Over the last few years many different quantifiers of asymmetry and coherence have been studied. Examples include Wigner-Yanase skew information [30, 33, 65, 73], Fisher information [65–67], functions defined based on the notion of modes of asymmetry [74], robustness of asymmetry [75], the relative entropy of asymmetry [41, 76] and its generalization, called Holevo measures of asymmetry [65]. Each of these functions provides a different way of quantifying the resource of coherence, and their monotonicity yield necessary conditions on state transformations. All the aforementioned functions, however, fail to see a simple, yet fundamental feature of quantum coherence: Given any finite copies of a generic mixed coherent state, it is impossible to generate a single copy of a *pure* coherent state, with a non-zero probability.

Here, we introduce a new measure of asymmetry which captures the missing part of the picture and predicts the unreachability of pure coherent states. For a system with state ρ let the *Purity of Coherence* with respect to the eigenbasis of

an observable H be

$$P_H(\rho) \equiv \text{Tr}(H\rho^2 H \rho^{-1}) - \text{Tr}(\rho H^2)$$
 (9a)

$$= \sum_{j,k} \frac{p_k^2 - p_j^2}{p_j} |\langle \psi_k | H | \psi_j \rangle|^2 , \qquad (9b)$$

if $\sup(H\rho H)\subseteq\sup(\rho)$, and $P_H(\rho)=\infty$ otherwise, where $\rho=\sum_j p_j|\psi_j\rangle\langle\psi_j|$ is the spectral decomposition of ρ . Equivalently, $P_H(\rho)=\infty$ if $[\Pi_\rho,H]\neq 0$, where Π_ρ is the projector to the support of ρ .

It turns out that this function has all the properties (**i-iv**) listed above for QFI, and therefore is a faithful, additive, convex measure of asymmetry. Similar to QFI, function $P_H(\rho)$ also determines how fast state ρ becomes distinguishable from its time evolved version $e^{-iHt}\rho e^{iHt}$ and is related to the (Petz) relative Renyi entropies (See the SM) [77, 78].

Despite all these similarities, QFI and purity of coherence have drastically different behaviors on pure states: QFI reduces to the energy variance, whereas the purity of coherence is ∞ , unless the state is an energy eigenstate, in which case it is zero (Note that any additive measure of asymmetry which can predict the unreachability of pure states from a generic mixed state, should be ∞ on pure states). The unboundedness of the purity of coherence reflects the fact that, in some sense, the coherence content of a single qubit can be arbitrarily large. As we discuss later, the fact that there are two additive measures of resource with completely different behaviors has deep consequences.

Combining the above properties of the purity of coherence, one can easily show that if there is a TI operation which transforms n copies of state ρ to state σ , with probability p, then $n \geq pP_H(\sigma)/P_H(\rho)$. Therefore, to obtain a pure coherent state σ with non-zero probability of success $p \neq 0$, one needs $P_H(\rho) = \infty$ or $n = \infty$.

Remark 6. Similar to all other measures of asymmetry, both functions QFI and the purity of coherence are non-increasing under thermal operations, and generalized thermal processes introduced in [79].

In the context of quantum thermodynamics, the monotonicity of each of these measures of asymmetry under thermal operations can be interpreted as an extension of the second law of thermodynamics [1–4].

A. Relation with Quantum Fisher Information

A closer look at the properties of the purity of coherence reveals an interesting relation with QFI. Comparing Eq.(9b) with Eq.(4) one can show that the purity of coherence is always larger than or equal to QFI, i.e.

$$P_H(\rho) > F_H(\rho) . \tag{10}$$

Furthermore, calculating the purity of coherence for qubits, one finds the nice formula

$$P_H(\rho) = \frac{F_H(\rho)}{2[1 - \text{Tr}(\rho^2)]}$$
 (11)

Hence, the purity of coherence is determined by a combination of QFI and the purity of state. For states close to the totally mixed state, one finds $2[1-{\rm Tr}(\rho^2)]\approx 1$, which implies the purity of coherence is approximately equal to QFI. On the other hand, as state ρ becomes closer to a pure state, $P_H(\rho)$ becomes arbitrary larger than QFI.

It turns out that this is a general feature of the purity of coherence, which holds beyond qubit systems. If ρ is ϵ -close to the totally mixed state I/d in the trace distance, then

$$\frac{P_H(\rho)}{F_H(\rho)} = 1 + \mathcal{O}(\epsilon) \quad \text{if} \quad \|\rho - \frac{I}{d}\|_1 \le \epsilon \ll 1. \quad (12)$$

Remarkably, in this limit the dominant behavior of the purity of coherence only depends on QFI and their ratio is approximately one. In the opposite limit, on the other hand, where state ρ is close to a pure state, we find

$$P_H(\rho) \ge \frac{F_H(\psi_{\text{max}})}{4} \times \left[\frac{p_{\text{max}}^2}{1 - p_{\text{max}}} - 1\right],$$
 (13)

where $p_{\rm max}$ is the largest eigenvalue of ρ , and $\psi_{\rm max}$ is the corresponding eigenvector. Again, as the state becomes more pure both $p_{\rm max}$ and the purity ${\rm Tr}(\rho^2)$ converge to one, which imply $P_H(\rho)$ diverges, unless $\psi_{\rm max}$ is an energy eigenstate.

Based on this result together with Eq.(11) we conclude that, roughly speaking, the purity of coherence is lower bounded by the ratio of QFI (for a pure state close to ρ) to one minus the purity of state. In other words, higher $P_H(\rho)$ means more pure coherence, which justifies its name, the purity of coherence.

B. Analogy with the total and free energies

In the previous section we saw that for any mixed state ρ of system S we can find a pure state $|\Phi_{\rho}\rangle_{SA}$ which purifies ρ , and has the same QFI. Therefore, if we start from $|\Phi_{\rho}\rangle_{SA}$ and discard the system A, the QFI does not decrease. Hence, one may expect that the process is reversible, that is there exists a TI operation which transforms state ρ back to $|\Phi_{\rho}\rangle_{SA}$ (Note that discarding the subsystem A is itself a TI operation). However, using the purity of coherence we find that this process is indeed irreversible; by discarding system A we loose purity, and consequently the purity of coherence decreases, even though QFI remains constant. Since the purity of coherence is monotone under TI operations, we conclude that the process is irreversible. From the point of view of quantum thermodynamics, this means that, even by spending an arbitrary amount of work, we cannot recover the original pure state $|\Phi_{\rho}\rangle_{SA}$.

This example along with other properties of the purity of coherence that we saw previously, suggest that the relation between the purity of coherence and QFI is analogous to the relation between the free and total energies in thermodynamics.

The latter functions, which quantify the amount of thermodynamics resources, are both additive and yet are independent of each other. In particular, the free energy distinguishes ordered (low-entropy) energy and disordered (high-entropy) energy. Similarly, QFI and the purity of coherence are both additive quantifiers of coherence and are each relevant in some contexts. It turns out that for some applications the same amount of coherence quantified by QFI in states with more purity is a more useful resource, and the purity of coherence can recognize the distinction between the pure and mixed coherence. An important example is coherence distillation.

V. COHERENCE DISTILLATION

We saw that the purity of coherence distinguishes between coherence in mixed and pure states. This makes it a powerful tool for studying coherence distillation process in both single-shot and asymptotic regimes. In particular, using the purity of coherence we find a tight bound on the minimum achievable error for distillation in the single-copy regime, which suggests that this quantity is properly characterizing the unreachability of pure coherent states.

A. Asymptotic regime

The distillable coherence of a system with Hamiltonian H and state ρ with period τ relative to a standard coherent pure state $\Phi_{\text{c-bit}}$ with the same period, is the maximum rate at which $\Phi_{\text{c-bit}}$ can be obtained from ρ using TI operations.

Definition 7. The distillable coherence (w.r.t. TI operations) of a system with state ρ and Hamiltonian H is defined as

$$C_d^{T\!I}(\rho) = \sup R: \rho^{\otimes n} \xrightarrow{TI} \stackrel{\epsilon}{\approx} \Phi_{c\text{-}bit}^{\otimes \lceil Rn \rceil} \text{ as } n \to \infty, \epsilon \to 0 \; .$$

The main result of this section is the following theorem.

Theorem 8. For any state with bounded purity of coherence the distillable coherence (w.r.t. TI operations) is zero. For bounded Hamiltonians, $P_H(\rho) < \infty$ iff $[\Pi_\rho, H] = 0$, where Π_ρ is the projector to the support of ρ . Therefore,

$$[\Pi_{\rho}, H] = 0 \iff P_H(\rho) < \infty \implies C_d^{TI}(\rho) = 0.$$
 (14)

We find that, surprisingly, for a generic mixed state, which has a full-rank density operator, $[\Pi_{\rho}, H] = 0$ and therefore the distillable coherence is zero! In other words, a typical coherent state is a *bound* resource [80], meaning that no pure state resource can be distilled from it, but in order to create it, nonzero amount of pure state resource is required. In the context of quantum thermodynamics this means that, even if we spend an unbounded amount of work, a thermal operation cannot distill pure coherence at a nonzero rate from a generic mixed state. Hence the hypothetical coherence distillation machine depicted in Fig.(1) is impossible.

In the rest of this section we sketch the proof of theorem 8 (See the SM for details). Suppose there exists a TI oper-

ation which transforms $\rho^{\otimes n}$ to a state whose distance to the desired pure state $\Phi^{\otimes \lceil Rn \rceil}_{\text{c-bit}}$ vanishes in the limit n goes to infinity, and consider the purity of coherence of such an output state. As we saw in the previous section, roughly speaking, the purity of coherence is lower bounded by the ratio of QFI to one minus the purity of state. As n grows and the output state converges to the desired state $\Phi_{\text{c-bit}}^{\otimes \lceil Rn \rceil}$ its QFI increases linearly with Rn. But, at the same time, its purity also increases; in the limit the state is completely pure. This means that to converge to the ideal output state $\Phi_{\text{c-bit}}^{\otimes \lceil Rn \rceil}$ for R>0, the purity of coherence of the output should grow faster than linear in n. On the other hand, using the additivity of purity of coherence, we find that the total purity of coherence at the input is $nP_H(\rho)$, which grows linearly with n. But the purity of coherence is non-increasing under TI operations. This means that either $P_H(\rho) = \infty$ or R = 0, which proves theorem 8. In the SM we present a rigorous version of this argument which holds even if we use a *helper* pure state χ at the input, provided that its Hamiltonian is bounded and its Hilbert space is finite-dimensional (In this case the purity of coherence of the input $\rho^{\otimes n} \otimes \chi$ is ∞ , and the argument becomes more complicated).

It is interesting to note that if instead of the purity of coherence, we just look at the QFI of the input and output, the impossibility of distillation cannot be seen, because unlike the purity of coherence, QFI grows linearly for both the input and the desired output states.

B. Single-shot regime

Coherence distillation is also interesting in the single-shot regime: Having n copies of state ρ as the input, how close can we get to a desired pure state ψ , using only TI operations? As we explain in the SM, using the results of [79], the maximum achievable fidelity is given by a simple formula, $\max_{\mathcal{E}_{\text{TI}}} \langle \psi | \mathcal{E}_{\text{TI}}(\rho^{\otimes n}) | \psi \rangle = 2^{-H_{\min}(B|A)}$, where the maximization is over all TI operations, and H_{\min} is the *conditional minentropy* [78, 81, 82] for a particular bipartite state constructed from ψ and $\rho^{\otimes n}$ (See the SM).

Although interesting, such a formula does not clearly show how large should n be to achieve a certain desired error. On the other hand, using the purity of coherence one finds a simple lower bound on n, which is tight in the large n limit.

1. Example: single-shot coherence distillation for qubits

The smallest quantum clock is a qubit with two different energy levels (See e.g. [24, 25, 83]). Suppose we are interested to prepare a qubit clock with Hamiltonian $H_{\text{c-bit}} = \pi \sigma_z / \tau$ in a state close to $|\Phi\rangle_{\text{c-bit}} = (|0\rangle + |1\rangle)/\sqrt{2}$. We start from n copies of the system in a noisy version of state $|\Phi\rangle_{\text{c-bit}}$, i.e. in state $\rho = \lambda |\Phi\rangle\langle\Phi|_{\text{c-bit}} + (1-\lambda)I/2$ with $0 < \lambda < 1$, and want to distill one copy with higher fidelity. What is the maximum achievable fidelity $\langle\Phi|\mathcal{E}_{\text{TI}}(\rho^{\otimes n})|\Phi\rangle_{\text{c-bit}}$ using a TI operation \mathcal{E}_{TI} ?

Using the monotonicity of QFI one can find a simple bound on n: If $\rho^{\otimes n} \xrightarrow{TI} \sigma$, then $n \geq F_H(\sigma)/F_H(\rho)$. But, because Fisher information is bounded by the energy variance, as $F_H(\rho) \leq 4V_H(\rho)$, for bounded Hamiltonians the QFI is always finite. This means that this bound can always be satisfied for a finite n, even if we choose $\sigma = |\Phi\rangle\langle\Phi|_{\text{c-bit}}$. Using the purity of coherence, on the other hand, and combining the bound $n \geq P_H(\sigma)/P_H(\rho)$ with the formula for the purity of coherence in Eq.(11) we can show that: In the limit of large n, for any TI operation \mathcal{E}_{TI} the *infidelity* of state $\mathcal{E}_{\text{TI}}(\rho^{\otimes n})$ with the desired pure state $\Phi_{\text{c-bit}}$ is lower bounded by

$$1 - \langle \Phi | \mathcal{E}_{\text{TI}}(\rho^{\otimes n}) | \Phi \rangle_{\text{c-bit}} \ge \frac{1}{n} \frac{1 - \lambda^2}{4\lambda^2} + \mathcal{O}(\frac{1}{n^2}) \,. \tag{15}$$

Remarkably, this bound is tight (up to a factor of 2)! Cirac et al. [84] have studied a related problem in a paper titled "Optimal purification of single qubits", and have shown that there exists a quantum operation $\mathcal{E}_{\text{Schur}}$ (related to the Schur transformation) which is covariant with respect to the full unitary group SU(2), for which the infidelity $1-\langle\Phi|\mathcal{E}_{\text{Schur}}(\rho^{\otimes n})|\Phi\rangle_{\text{c-bit}}$ is equal to $2/(1+\lambda)$ times the right-hand side of this bound. But, since this operation is covariant with respect to the full unitary group, it is also covariant with respect to time translations, i.e. satisfies Eq.(1).

Therefore, this bound, which is dictated by the monotonicity of the purity of coherence, determines the lowest possible infidelity $1-\max_{\mathcal{E}_\Pi} \langle \psi | \mathcal{E}_{\Pi}(\rho^{\otimes n}) | \psi \rangle$, up to a factor $2/(1+\lambda)$, which approaches 1 for input states close to $|\Phi\rangle_{\text{c-bit}}$ (i.e. for $\lambda \approx 1$) and never exceeds 2. It remains an open question if using TI operations, which are not necessarily covariant with respect to the full SU(2) group, one can achieve this bound for a general λ .

VI. DISCUSSION

In the recent years we have seen a significant progress in understanding the concept of coherence in the framework of quantum resource theories and, in particular, in the context of quantum thermodynamics (See e.g. [1–4, 7, 59, 85, 86]). In spite of this progress, some aspects of coherence are not well-understood yet. Here, we have highlighted an important feature of quantum coherence which manifests itself, for instance, in the unreachability of pure coherent states from mixed states in both the single and asymptotic regimes, and the fact that (in some precise sense) the coherence content of a single qubit can be arbitrarily large.

To quantify this feature of coherence, we introduced a new additive measure of asymmetry, called the purity of coherence and showed that the monotonicity of this quantity under TI operations gives a tight bound on the coherence distillation in the single-shot regime. This observation supports the idea that the purity of coherence is adequately quantifying this feature of coherence. We also used the purity of coherence to study the coherence distillation process in the asymptotic regime, and found that the distillable coherence for a generic mixed state is zero. In other words, the hypothetical coherence dis-

tillation machine depicted in Fig.(1), which purifies quantum clocks (or equivalently, distills coherence in the energy eigenbasis) is not consistent with the laws of quantum mechanics and thermodynamics.

Investigating the manipulation of coherence also led us to new properties of QFI (i.e. theorems 2 and 5), which is arguably the most fundamental and the most studied quantity of quantum estimation theory. In particular, we found a novel operational interpretation of QFI in terms of the coherence cost of preparing a general mixed state from a standard pure state. QFI has found extensive applications in different areas of physics, such as quantum information theory [87–90], quantum speed limits [54, 91, 92], many-body systems [93–96], and quantum gravity [97]. Finding possible implications of theorems 2 and 5 in these areas can be interesting.

The fact that in thermodynamics entropy and energy (or equivalently, free and total energies) are two independent additive state functions, have deep consequences. The impossibility of perpetual motion machines, for instance, cannot be seen if one only considers energy and not entropy. The same is expected to be true for QFI and the purity of coherence. Here, we saw one of these consequences, namely the impossibility of coherence distillation from a generic mixed state. It is worth noting that there are infinitely many other examples of additive measures of asymmetry. But, in most cases their behaviors are not independent of each other. For instance, the Wigner-Yanase skew information [33], $W_H(\rho) = -\text{Tr}([\sqrt{\rho}, H]^2)/2$, is another measure of asymmetry [30, 65, 73], which satisfies all properties (i-iv). However, the behavior of skew information is predictable from QFI, because $F_H(\rho)/2 \leq W_H(\rho) \leq F_H(\rho)$ [62]. The question of classifying all independent additive measures of asymmetry remains open.

Another important remaining open question is extending theorem 1, which determines the necessary and sufficient condition for asymptotic transformations between pure states, as well as extending its generalization in theorem 5, which allows mixed states as the output, to the case of operations which are covariant under a general symmetry group, such as SO(3). For a general group, one expects additional constraints beyond conservation of QFI [67] (we note that recently there has been an important progress in generalizing theorem 1 for group SO(3) and a special class of pure states [98]).

General symmetry groups are also relevant in the context of quantum thermodynamics. In particular, if in addition to the Hamiltonian, there are other thermodynamic conserved additive observables, such as angular momentum, then state transformations should be covariant with respect to a larger symmetry group, which, in general, is non-commutative [99–101]. In this case, using measures of asymmetry for the larger symmetry group one can find additional constraints [100]. For instance, if the total angular momentum operator L_z of the system-bath is conserved, then both functions F_{L_z} and P_{L_z} are also non-increasing in any state transformation. The monotonicity of these functions can be interpreted as extensions of the second law of thermodynamics.

In this paper, we focused on the implications of our results in the context of quantum clocks and thermodynamics. Another important application of the resource theory of asymmetry is in the context of metrology [53, 65]. Finding the implications of our results in this context is left for future works. **Acknowledgments:** I am grateful to David Jennings for read-

ing the manuscript carefully, and providing many useful comments and suggestions.

- M. Lostaglio, D. Jennings, and T. Rudolph, Nature communications 6 (2015).
- [2] M. Lostaglio, K. Korzekwa, D. Jennings, and T. Rudolph, Physical Review X 5, 021001 (2015).
- [3] K. Korzekwa, M. Lostaglio, J. Oppenheim, and D. Jennings, New Journal of Physics 18, 023045 (2016).
- [4] V. Narasimhachar and G. Gour, Nature communications 6, 7689 (2015).
- [5] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. 81, 865 (2009).
- [6] A. Winter and D. Yang, Physical review letters 116, 120404 (2016).
- [7] A. Streltsov, G. Adesso, and M. B. Plenio, Reviews of Modern Physics 89, 041003 (2017).
- [8] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
- [9] A. S. Holevo, Probabilistic and Statistical Aspects of Quantum Theory (North-Holland, Amsterdam, 1982).
- [10] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. 72, 3439 (1994).
- [11] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. 96, 010401 (2005).
- [12] V. Giovannetti, S. Lloyd, and L. Maccone, Nature Photonics 5, 222 (2011).
- [13] M. Horodecki and J. Oppenheim, International Journal of Modern Physics B 27, 1345019 (2013).
- [14] F. G. Brandão and G. Gour, Physical review letters 115, 070503 (2015).
- [15] B. Coecke, T. Fritz, and R. W. Spekkens, Information and Computation 250, 59 (2016).
- [16] W. K. Wootters and W. H. Zurek, Nature 299, 802 (1982).
- [17] D. Dieks, Physics Letters A 92, 271 (1982).
- [18] H. Salecker and E. Wigner, Physical Review **109**, 571 (1958).
- [19] A. Peres, American Journal of Physics 48, 552 (1980).
- [20] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Reviews of Modern Physics 79, 555 (2007).
- [21] V. Giovannetti, S. Lloyd, L. Maccone, and F. Wong, Physical review letters 87, 117902 (2001).
- [22] V. Giovannetti, S. Lloyd, and L. Maccone, Nature 412, 417 (2001).
- [23] V. Giovannetti, S. Lloyd, L. Maccone, and M. Shahriar, Physical Review A 65, 062319 (2002).
- [24] V. Bužek, R. Derka, and S. Massar, Physical review letters 82, 2207 (1999).
- [25] G. Chiribella, Y. Yang, and A. C.-C. Yao, Nature communications 4, 2915 (2013).
- [26] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Y. Halpern, Physics Reports 583, 1 (2015).
- [27] G. Gour, Physical Review A 95, 062314 (2017).
- [28] G. Gour and R. W. Spekkens, New Journal of Physics 10, 033023 (2008).
- [29] I. Marvian and R. W. Spekkens, New Journal of Physics 15, 033001 (2013).
- [30] I. Marvian, Ph.D. thesis, University of Waterloo, https://uwspace.uwaterloo.ca/handle/10012/7088 (2012).

- [31] M. Nielsen and I. Chuang, Quantum Computation and Quantum Information, Cambridge Series on Information and the Natural Sciences (Cambridge University Press, 2000), ISBN 9780521635035.
- [32] M. M. Wilde, Quantum information theory (Cambridge University Press, 2013).
- [33] E. P. Wigner and M. M. Yanase, Proceedings of the National Academy of Sciences of the United States of America 49, 910 (1963).
- [34] H. Araki and M. M. Yanase, Physical Review 120, 622 (1960).
- [35] Y. Aharonov and L. Susskind, Physical Review 155, 1428 (1967).
- [36] A. Peres and P. F. Scudo, arXiv preprint quant-ph/0201017 (2002).
- [37] N. Gisin and S. Popescu, Phys. Rev. Lett. 83, 432 (1999).
- [38] A. Kitaev, D. Mayers, and J. Preskill, Physical Review A 69, 052326 (2004).
- [39] N. Schuch, F. Verstraete, and J. I. Cirac, Physical review letters 92, 087904 (2004).
- [40] N. Schuch, F. Verstraete, and J. I. Cirac, Physical Review A 70, 042310 (2004).
- [41] J. A. Vaccaro, F. Anselmi, H. M. Wiseman, and K. Jacobs, Physical Review A 77, 032114 (2008).
- [42] G. Chiribella, G. D'ariano, and M. Sacchi, Physical Review A 72, 042338 (2005).
- [43] M. Skotiniotis and G. Gour, New Journal of Physics 14, 073022 (2012).
- [44] S. D. Bartlett, T. Rudolph, R. W. Spekkens, and P. S. Turner, New J. Phys. 8, 58 (2006).
- [45] M. Ahmadi, D. Jennings, and T. Rudolph, New Journal of Physics 15, 013057 (2013).
- [46] I. Marvian and R. W. Spekkens, arXiv preprint arXiv:1212.3378 (2012).
- [47] P. Hayden, S. Nezami, S. Popescu, and G. Salton, arXiv preprint arXiv:1709.04471 (2017).
- [48] R. M. Angelo, N. Brunner, S. Popescu, A. J. Short, and P. Skrzypczyk, Journal of Physics A: Mathematical and Theoretical 44, 145304 (2011).
- [49] I. Marvian and R. Mann, Physical Review A 78, 022304 (2008).
- [50] J. Eisert, T. Felbinger, P. Papadopoulos, M. Plenio, and M. Wilkens, Physical review letters 84, 1611 (2000).
- [51] J. Å berg, Physical review letters 113, 150402 (2014), ISSN 1079-7114.
- [52] I. Marvian and S. Lloyd, arXiv preprint arXiv:1608.07325 (2016).
- [53] I. Marvian and R. W. Spekkens, Physical Review A 94, 052324 (2016).
- [54] I. Marvian, R. W. Spekkens, and P. Zanardi, Physical Review A 93, 052331 (2016).
- [55] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and T. Beth, Int. J. Theor. Phys. 39, 2717 (2000).
- [56] M. Horodecki and J. Oppenheim, Nat. Commun. 4, 1 (2013).
- [57] F. G. Brandao, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, Physical review letters 111, 250404 (2013).

- [58] J. Åberg, Nature communications 4, 1925 (2013).
- [59] J. Goold, M. Huber, A. Riera, L. del Rio, and P. Skrzypczyk, Journal of Physics A: Mathematical and Theoretical 49, 143001 (2016).
- [60] N. H. Y. Ng, L. Mančinska, C. Cirstoiu, J. Eisert, and S. Wehner, New Journal of Physics 17, 085004 (2015).
- [61] Theorem 25 in [30] (2012).
- [62] D. Petz and C. Ghinea, in *Quantum probability and related topics* (World Scientific, 2011), pp. 261–281.
- [63] M. G. Paris, International Journal of Quantum Information 7, 125 (2009).
- [64] M. Hayashi, Quantum Information (Springer, 2006).
- [65] I. Marvian and R. W. Spekkens, Nature communications 5, 3821 (2014).
- [66] B. Yadin and V. Vedral, Physical Review A 93, 022122 (2016).
- [67] I. Marvian and R. W. Spekkens, Physical Review A 90, 014102 (2014).
- [68] H. Kwon, H. Jeong, D. Jennings, B. Yadin, and M. Kim, arXiv preprint arXiv:1711.03395 (2017).
- [69] G. Chiribella, arXiv preprint arXiv:1804.01943 (2018).
- [70] G. Tóth and D. Petz, Physical Review A 87, 032324 (2013).
- [71] S. Yu, arXiv preprint arXiv:1302.5311 (2013).
- [72] B. Toloui, G. Gour, and B. C. Sanders, Physical Review A 84, 022322 (2011).
- [73] D. Girolami, Physical review letters 113, 170401 (2014).
- [74] I. Marvian and R. W. Spekkens, Physical Review A 90, 062110 (2014).
- [75] M. Piani, M. Cianciaruso, T. R. Bromley, C. Napoli, N. Johnston, and G. Adesso, arXiv preprint arXiv:1601.03782 (2016).
- [76] G. Gour, I. Marvian, and R. W. Spekkens, Physical Review A 80, 012307 (2009).
- [77] D. Petz, Reports on mathematical physics 23, 57 (1986).
- [78] M. Tomamichel, Quantum Information Processing with Finite Resources: Mathematical Foundations, vol. 5 (Springer, 2015)
- [79] G. Gour, D. Jennings, F. Buscemi, R. Duan, and I. Marvian, arXiv preprint arXiv:1708.04302 (2017).
- [80] M. Horodecki, P. Horodecki, and R. Horodecki, Physical Review Letters 80, 5239 (1998).
- [81] R. Renner, International Journal of Quantum Information **6**, 1
- [82] R. Konig, R. Renner, and C. Schaffner, IEEE T. Inform. Theory 55, 4337 (2009).
- [83] M. P. Woods, R. Silva, and J. Oppenheim, arXiv preprint arXiv:1607.04591 (2016).
- [84] J. Cirac, A. Ekert, and C. Macchiavello, Physical review letters 82, 4344 (1999).
- [85] G. Chiribella and Y. Yang, Physical Review A 96, 022327

- (2017).
- [86] N. Y. Halpern and J. M. Renes, Physical Review E 93, 022126 (2016).
- [87] D. Girolami, T. Tufarelli, and G. Adesso, Physical review letters 110, 240402 (2013).
- [88] S. Kim, L. Li, A. Kumar, and J. Wu, Physical Review A 97, 032326 (2018).
- [89] C. Zhang, B. Yadin, Z.-B. Hou, H. Cao, B.-H. Liu, Y.-F. Huang, R. Maity, V. Vedral, C.-F. Li, G.-C. Guo, et al., Physical Review A 96, 042327 (2017).
- [90] D. Girolami and B. Yadin, Entropy 19, 124 (2017).
- [91] D. P. Pires, M. Cianciaruso, L. C. Céleri, G. Adesso, and D. O. Soares-Pinto, Physical Review X 6, 021031 (2016).
- [92] D. Mondal, C. Datta, and S. Sazim, Physics Letters A 380, 689 (2016).
- [93] P. Zanardi, P. Giorda, and M. Cozzini, Physical review letters 99, 100603 (2007).
- [94] P. Zanardi, M. G. Paris, and L. C. Venuti, Physical Review A 78, 042105 (2008).
- [95] P. Zanardi, L. Campos Venuti, and P. Giorda, Physical Review A 76, 062318 (2007).
- [96] L. Campos Venuti and P. Zanardi, Physical Review Letters 99, 095701 (2007).
- [97] N. Lashkari and M. Van Raamsdonk, Journal of High Energy Physics 2016, 153 (2016).
- [98] Y. Yang, G. Chiribella, and Q. Hu, New Journal of Physics 19, 123003 (2017).
- [99] N. Y. Halpern, P. Faist, J. Oppenheim, and A. Winter, Nature communications 7, 12051 (2016).
- [100] M. Lostaglio, D. Jennings, and T. Rudolph, New Journal of Physics 19, 043008 (2017).
- [101] Y. Guryanova, S. Popescu, A. J. Short, R. Silva, and P. Skrzypczyk, Nature communications 7, ncomms12049 (2016).
- [102] A. D. Barbour, V. Ćekanavićius, et al., The Annals of Probability 30, 509 (2002).
- [103] A. Röllin and N. Ross, Bernoulli **21**, 851 (2015).
- [104] J. A. Adell and P. Jodrá, Journal of Inequalities and Applications 2006, 64307 (2006).
- [105] C. A. Fuchs and J. Van De Graaf, IEEE Transactions on Information Theory 45, 1216 (1999).
- [106] P. M. Hayden, M. Horodecki, and B. M. Terhal, Journal of Physics A: Mathematical and General 34, 6891 (2001).
- [107] Z.-W. Liu, X. Hu, and S. Lloyd, Physical review letters 118, 060502 (2017).
- [108] T. Rudolph, R. W. Spekkens, and P. S. Turner, Physical Review A 68, 010301 (2003).

Supplementary Material:

Coherence distillation machines are impossible in quantum thermodynamics

Contents

• Section A: Pure state transformations in the asymptotic (iid) regime

In this section we prove the first part of theorem 1 in the paper, which determines the rate of interconversion between pure states.

- Section B: Quantum Fisher Information: Preliminaries
- Section C: Quantum Fisher Information as the single-shot coherence cost

In this section we prove theorem 2 in the paper, and present a new proof of theorem 4.

• Section D: Monotonicity of Fisher information in approximate asymptotic transformations

In this section we prove that Fisher information cannot increase in the approximate asymptotic transformations. This implies that Quantum Fisher information is a lower bound on coherence cost. It also proves the second part of theorem 1 in the paper.

• Section E: Quantum Fisher Information as the coherence cost: iid regime

In this section we prove prove theorem 5 in the paper which states that the coherence cost is equal to Quantum Fisher Information.

• Section F: Purity of Coherence

In this section we introduce purity of coherence and study its properties. This section includes the following subsections:

- Connection with relative Renyi entropy
- Purity of coherence is lower-bounded by Quantum Fisher Information (Proof of Eq.(10))
- Purity of coherence for Qubits (Proof of Eq.(11))
- Purity of coherence for states close to the totally mixed state (Proof of Eq.(12))
- Purity of coherence for states close to pure states (Proof of Eq.(13))

• Section G: Distillable Coherence

In this section we prove theorem 8.

• Section H: Distillation in the single-shot regime

- In Sec.H 1 we present a simple formula for the maximum achievable fidelity of distillation, in terms of conditional min-entropy.
- In Sec.H 2 we prove Eq.(15) in the paper.

Appendix A: Pure state transformations in the asymptotic (iid) regime

In this section we study pure state transformations in the many-copy (iid) regime. In particular, we prove that the interconversion between pure states is possible with a rate less than or equal to the ratio of the energy variances of the input to the output. Then, in Sec.(D) we present theorem 22 which shows that the interconversion is not possible with a higher rate.

1. Single-copy pure state transformations

A fundamental fact about pure state conversions under TI operations is that the only relevant information about a pure state is encoded in the energy distribution of state. Let

$$H = \sum_{E} E \,\Pi_{E} \tag{A1}$$

be the spectral decomposition of the system Hamiltonian H, and ψ be the state of system. We assume the dynamic of ψ under Hamiltonian H is periodic with period τ , i.e.

$$\tau = \inf_{t} \{ t > 0 : |\langle \psi | e^{-iHt} | \psi \rangle| = 1 \} , \qquad (A2)$$

where we have assumed $\hbar=1$. This means that the set of energy levels E with nonzero probability, i.e. $\{E:\langle\psi|\Pi_E|\psi\rangle\neq0\}$, can be written as

$$E = \frac{2\pi n}{\tau} + E_0 \qquad n \in \mathbb{Z} \,, \tag{A3}$$

where E_0 defines the energy reference. By shifting the energy reference we can always choose $E_0 = 0$. In other words, we can always choose energy reference such that

$$e^{-iH\tau}|\psi\rangle = |\psi\rangle.$$
 (A4)

In the rest of this paper, we assume the energy references are chosen such that this condition is satisfied.

Next, we define the distribution p_{ψ} , as

$$p_{\psi}(n) = \langle \psi | \Pi_{2\pi n/\tau} | \psi \rangle \tag{A5}$$

$$= \frac{1}{2\pi} \int_0^{2\pi} d\theta \ e^{i\theta n} \ \langle \psi | e^{-iH\tau \frac{\theta}{2\pi}} | \psi \rangle \tag{A6}$$

if $2\pi n/\tau$ is an eigenvalue of H, and $p_{\psi}(n)=0$ otherwise. Therefore, the probability distribution p_{ψ} , which is defined over integers \mathbb{Z} , uniquely determines the energy distribution of state ψ : the energy of state is $2\pi n/\tau$ with probability $p_{\psi}(n)$. In the following we sometimes refer to p_{ψ} as the energy distribution of ψ .

Consider two different pure states ψ and ϕ of a system with Hamiltonian H. It can be shown [28, 29, 67] that ψ can be transformed to ϕ via an energy-conserving unitary V, i.e. $V|\psi\rangle=|\phi\rangle$ and [V,H]=0, if and only if they have the same energy distributions, i.e.

$$\forall n \in \mathbb{Z}: \quad p_{\psi}(n) = p_{\phi}(n) \,, \tag{A7}$$

or equivalently, if and only if, they have the same characteristic functions [29, 67], i.e.

$$\forall \theta \in (0, 2\pi] : \quad \langle \psi | e^{-iH\tau \frac{\theta}{2\pi}} | \psi \rangle = \langle \phi | e^{-iH\tau \frac{\theta}{2\pi}} | \phi \rangle . \tag{A8}$$

Therefore, for a given Hamiltonian H, if we know the probability distribution p_{ψ} we know all the relevant information about state ψ from the point of view of transformations under TI operations.

The above result can be generalized in the following sense: if the energy distributions p_{ψ} and p_{ϕ} are close to each other in the total variation distance (trace distance) then there exists a unitary transformation which transforms ψ to a state close to ϕ [29].

In particular, there exists a unitary V which commutes with the system Hamiltonian H such that (Theorem 3 in [29])

$$|\langle \phi | V | \psi \rangle| = \sum_{n} \sqrt{p_{\psi}(n)p_{\phi}(n)} \ge 1 - \frac{1}{2} \|p_{\psi} - p_{\phi}\|_{1},$$
 (A9)

$$\equiv 1 - d_{\text{TV}}(p_{\psi}, p_{\phi}) \,, \tag{A10}$$

where

$$d_{\text{TV}}(p_{\psi}, p_{\phi}) \equiv \frac{1}{2} \|p_{\psi} - p_{\phi}\|_{1} = \frac{1}{2} \sum_{n} |p_{\psi}(n) - p_{\phi}(n)|, \qquad (A11)$$

is the total variation distance (trace distance) between the two distributions.

In addition to the energy-conserving unitaries, TI operations include transformations which do not conserve the energy. In particular, consider two systems with Hamiltonian H_1 and H_2 and states ψ_1 and ψ_2 , respectively, and assume both states have period τ . Again we assume energy references are chosen such that

$$e^{-iH_{1,2}\tau}|\psi_{1,2}\rangle = |\psi_{1,2}\rangle$$
 (A12)

Let p_{ψ_1} and p_{ψ_2} be the energy distributions for these two states defined via Eq.(A5). It can be shown that if there exists an integer k such that

$$\forall n \in \mathbb{Z}: \quad p_{\psi_1}(n) = p_{\psi_2}(n+k) , \qquad (A13)$$

then there exists a TI operation which transforms state ψ_1 to ψ_2 , or vice verse [28, 29, 67]. In the special case where the input and output Hamiltonians are identical, i.e. $H_1 = H_2$ this operation basically adds (subtracts) energy $k2\pi/\tau$ to the system.

Again, if Eq.(A13) holds approximately, then the transformation between ψ_1 and ψ_2 can be implemented approximately via a TI operation, with an error which is determined by the total variation distance between the probability distribution $p_{\psi_1}(n)$ and $p_{\psi_2}(n+k)$.

The following proposition summarizes these results

Proposition 9. (based on [28, 29, 67]) Suppose two systems with Hamiltonian H_1 and H_2 and states ψ_1 and ψ_2 both have period τ . Furthermore, assume the energy reference for Hamiltonians H_1 and H_2 are chosen such that $e^{-iH_{1,2}\tau}|\psi_{1,2}\rangle=|\psi_{1,2}\rangle$. Let $p_{\psi_{1,2}}$ be the energy distributions for two pure states $\psi_{1,2}$, defined by

$$p_{\psi_{1,2}}(n) = \frac{1}{2\pi} \int_0^{2\pi} d\theta \ e^{i\theta n} \langle \psi_{1,2} | e^{-iH_{1,2}\tau \frac{\theta}{2\pi}} | \psi_{1,2} \rangle \ . \tag{A14}$$

Suppose by translating the distribution p_{ψ_1} with an integer k we obtain distribution p_{ψ_2} with the total error ϵ in the total variation distance, i.e.

$$\frac{1}{2} \sum_{n} |p_{\psi_1}(n) - p_{\psi_2}(n+k)| = \epsilon.$$
 (A15)

Then, there exists a TI operation \mathcal{E}_{TI} such that $\mathcal{E}_{TI}(|\psi_1\rangle\langle\psi_1|)$ is a pure state which satisfies

$$\langle \psi_2 | \mathcal{E}_{TI}(|\psi_1\rangle \langle \psi_1|) | \psi_2 \rangle \ge 1 - 2\epsilon \ .$$
 (A16)

2. Transformations in the iid regime

Suppose we are given m copies of a system with Hamiltonian H and state ψ , i.e. , we are given non-interacting systems with the joint state $\psi^{\otimes m}$ and the total (non-interacting) Hamiltonian $H_{\text{tot}} = \sum_{i=1}^m H^{(i)}$, where $H^{(i)} = I^{\otimes (i-1)} \otimes H \otimes I^{\otimes (m-i)}$. The total energy for these systems is the sum of the energy of the individuals. Therefore, for state $\psi^{\otimes m}$ the probability distribution over energy eigenspaces of H_{tot} is equal to the probability distribution for the random variable $n_{\text{tot}} = n_1 + \dots + n_m$, where each integer random variables n_k has the probability distribution p_{ψ} . Hence, the probability distribution of the total energy for state $\psi^{\otimes m}$ is given by the m-fold convolution of the probability distribution p_{ψ} , i.e.

$$p_{\psi \otimes m} = \underbrace{p_{\psi} * \dots * p_{\psi}}_{m \text{ times}}. \tag{A17}$$

Then, in the limit of large number of copies $m\gg 1$, the central limit theorem implies that the distribution of energy for $\psi^{\otimes m}$ converges to a Gaussian, or equivalently, a shifted Poisson distribution. But, any such distribution is determined by only two parameters, i.e. the variance and the mean. It follows that if the energy variance for two states $\psi^{\otimes m}$ and $\phi^{\otimes \lceil Rm \rceil}$ match approximately, then by adding or subtracting energy, which is a TI operation, we can shift the center of the distributions and overlap them. This is the main intuition in the arguments of [28, 39, 40].

Although this intuition is correct, there are some crucial details which require more careful analysis. Most importantly, the standard central limit theorems do not guarantee convergence in the total variation distance. And, in fact, this convergence does not happen generally. For instance, consider two states

$$|\phi\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$$
 and $|\psi\rangle = (|0\rangle + |2\rangle)/\sqrt{2}$,

with the system Hamiltonian $H=2\pi\tau^{-1}\sum_{k=0}^{\infty}k|k\rangle\langle k|$. Then, one can easily see that for any m, state $|\psi\rangle^{\otimes m}$ has support only on even integers, while state $|\phi\rangle^{\otimes m}$ has support both on even and odd integers. It follows that their (shifted) energy distributions do not converge to each other, even in the infinite copy regime.

To avoid these situations, [28, 39, 40] make an extra assumption that the probability distributions p_{ψ} and p_{ϕ} are *gapless*, meaning that each has support only on a connected interval of integers. As we discuss later this assumption is not necessary for the existence of asymptotic transformations. For example, many copies of state $|\psi\rangle = (|0\rangle + |2\rangle + |3\rangle)/\sqrt{3}$, which has a gapped spectrum, can be reversibly transformed to many copies of state $|\phi\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, with a non-zero rate.

But, even making the assumption that p_{ψ} and p_{ϕ} are both gapless, the arguments of [28, 39, 40] only prove the pointwise convergence of the (shifted) energy distributions for state $\psi^{\otimes m}$ and $\phi^{\otimes \lceil Rm \rceil}$. However, here we are interested in the convergence of $\psi^{\otimes m}$ and $\phi^{\otimes \lceil Rm \rceil}$ in the trace distance, and this notion of convergence can be guaranteed only if the (shifted) energy distributions converge in the total variation distance (See proposition 9). To prove this stronger notion of convergence, we need to use more advanced results on the limit theorems, which we review in the following section.

3. Local limit theorems and convergence in the total variation distance

In the following $Y \sim P(\sigma^2)$ means the integer-valued random variable Y has Poisson distribution with variance $\sigma \geq 0$, such that any integer $l \geq 0$ occurs with probability $e^{-\sigma}\sigma^l/l!$. The Poisson distribution is specified by only one parameter σ , which determines both the variance and the mean of the distribution. We are interested in the more general family of integer-valued distributions obtained by translating Poisson distributions with integers, such that the variance and mean can be independent of each other. However, by translating with an integer we can only change the mean of the distribution in a discrete fashion. It follows that using this family of distributions we cannot really achieve arbitrary mean and variance. Nevertheless, for any desired mean μ and variance σ we can find a translated Poisson distribution whose mean is exactly μ and its variance is close to σ , such that their difference is less than one.

For any given μ and $\sigma>0$, let $Z\sim TP(\mu,\sigma^2)$ be a random variable which satisfies $Z-s\sim P(\sigma^2+\gamma)$ where the shift $s:=\lfloor\mu-\sigma\rfloor$ is an integer, and $\gamma:=\mu-\sigma^2-\lfloor\mu-\sigma^2\rfloor$, satisfies $0\leq\gamma<1$. This means Z-s has Poisson distribution with variance $\sigma^2+\gamma$. It follows that the random variable $Z\sim TP(\mu,\sigma^2)$ has mean μ , i.e. $\mathbb{E} Z=\mu$, and its variance is $\mathbb{E} Z^2-(\mathbb{E} Z)^2=\sigma^2+\gamma$, which is between σ^2 and σ^2+1 .

In the following we see that, under certain conditions, sum of integer-valued random variables can be approximated by translated Poisson distributions. See [102] for further details and proofs (See also [103]).

Let $W = \sum_{i=1}^{m} X_i$ be sum of m independent integer-valued random variables X_i , with mean $\mu_i = \mathbb{E}X_i$ and variance $\sigma_i^2 = \text{Var}X_i$, and bounded third moment, i.e. $\mathbb{E}|X_i^3| < \infty$. Let $\mu = \mathbb{E}W := \sum_{i=1}^{m} \mu_i$, and $\sigma^2 := \sum_{i=1}^{m} \sigma_i^2$ be the variance of W. Let $\mathcal{L}(X_i)$ be the distribution of the random variable X_i . Let

$$W_i = W - X_i = \sum_{j \neq i} X_j , \qquad (A18)$$

i.e. the sum of all the random variables except X_i . Let

$$d = \max_{1 \le i \le m} d_{\text{TV}}(\mathcal{L}(W_i), \mathcal{L}(W_i + 1)), \tag{A19}$$

be the total variation distance between the random variable W_i and its translated version $W_i + 1$. Note that in the limit m goes to infinity, if the distribution converges to a translated Poisson distribution, or a discrete normal distribution, then d goes to zero. Finally, define

$$\phi_i := \mathbb{E}\{X_i(X_i - 1)\} + \frac{|\mu_i - \sigma_i^2|}{\sigma_i^2} \,\mathbb{E}\{(X_i - 1)(X_i - 2)\} + \frac{1}{\sigma_i^2} \mathbb{E}|X_i(X_i - 1)(X_i - 2)| \,. \tag{A20}$$

Note that if $\sigma_i^2 > 0$ and the third moment $\mathbb{E}X_i^3$ is finite, then $|\phi_i|$ is also a finite number.

Theorem 10. (Theorem 3.1 in Barbour-Cekanavicius [102]) The total variation distance of the distribution of W and the translated Poisson distribution $TP(\mu, \sigma^2)$ is bounded by

$$d_{TV}(\mathcal{L}(W), TP(\mu, \sigma^2)) = \frac{1}{2} \|\mathcal{L}(W) - TP(\mu, \sigma^2)\|_1 \le \frac{2 + d\sum_{i=1}^m \phi_i}{\sigma^2}.$$
 (A21)

In the special case where all the random variables X_i have identical distributions with finite third moment, then in the limit of large m, the upper bound converges to a constant times d. Then, if d goes to zero, the distribution of random variable $W = \sum_{i=1}^{m} X_i$ converges to a translated Poisson distribution.

In general, it turns out that in the limit $m \to \infty$, d goes to zero, if for a nonzero fraction of the random variables X_i , the distribution of X_i and its translated version $X_i + 1$ have a nonzero overlap, such that

$$d_{\text{TV}}(\mathcal{L}(X_i), \mathcal{L}(X_i+1)) < 1. \tag{A22}$$

The following corollary of the theorem focuses on this special case. Let

$$\nu_i = \min\{\frac{1}{2}, 1 - d_{\text{TV}}(\mathcal{L}(X_i), \mathcal{L}(X_i + 1))\}.$$
(A23)

Note that $\nu_i = 0$ if and only if X_i and $X_i + 1$ have disjoint supports (This is the case, for instance, if X_i is nonzero only for even integers).

Corollary 11. (Corollary 3.2 in Barbour-Cekanavicius[102]) Let $a = \min_i \sigma_i$, and $b = \min_i \nu_i$, where $\nu_i = \min\{\frac{1}{2}, 1 - d_{TV}(\mathcal{L}(X_i), \mathcal{L}(X_i+1))\}$. Assume both a, b > 0. Let $c = \max_i \psi_i/\sigma_i^2$ and assume $c < \infty$. Then, the total variation distance of the distribution of W and the translated Poisson distribution $TP(\mu, \sigma^2)$ is bounded by

$$d_{TV}(\mathcal{L}(W), TP(\mu, \sigma^2)) = \frac{1}{2} \|\mathcal{L}(W) - TP(\mu, \sigma^2)\|_1 \le \frac{c}{\sqrt{mb - 1/2}} + \frac{2}{ma}.$$
 (A24)

It follows that if all these random variables have positive variances and $\nu_i > 0$, i.e. they are not perfectly distinguishable from their translated versions, then the sum $W = \sum_i X_i$ converges to a translated Poisson distribution.

We end this section by recalling another useful result on the total variation distance between Poisson distributions (See []).

Lemma 12. [104] The total variation distance between two Poisson distributions with variances $\sigma^2 + x$ and σ^2 , for $x \ge 0$, is bounded by

$$d_{TV}(P(\sigma^2), P(\sigma^2 + x)) = \frac{1}{2} ||P(\sigma^2) - P(\sigma^2 + x)||_1$$
(A25)

$$= \frac{1}{2} \sum_{n} \left| \frac{e^{-\sigma} \sigma^{n}}{n!} - \frac{e^{-\sqrt{\sigma^{2} + x}} \sqrt{(\sigma^{2} + x)^{n}}}{n!} \right|$$
 (A26)

$$\leq \min\{x, \sqrt{\frac{2}{e}}(\sqrt{\sigma^2 + x} - \sigma)\}. \tag{A27}$$

Therefore, for variance $\sigma^2 \ge 0$, we find that the total variation distance between $P(\sigma^2)$ and $P(\sigma^2 + x)$ is bounded by the ratio of the difference between the two variances to the square root of the variance, i.e.

$$d_{\text{TV}}(P(\sigma^2), P(\sigma^2 + x)) \le \frac{x}{\sigma}.$$
(A28)

4. Energy-distribution in the iid regime: Translated Poisson distribution

Next, we use these results to study the conversion of pure states in the many-copy regime using TI operations. As we saw before, for m copies of a system with state ψ and Hamiltonian H, the total energy distribution $p_{\psi^{\otimes m}}$ is determined by the random variable

$$n_{\text{tot}} = n_1 + \dots + n_m \,, \tag{A29}$$

where each n_i has distribution p_{ψ} . Applying corollary 11 we find that, provided that certain conditions (listed below) are satisfied, this distribution can be approximated by a translated Poisson distribution $TP(\mu, \sigma^2)$ with the mean

$$\mu = \mathbb{E}\{n_{\text{tot}}\} = m \times \mathbb{E}\{n\} = m \times \frac{\tau}{2\pi} \langle \psi | H | \psi \rangle , \qquad (A30)$$

and the variance

$$\sigma^2 = m \times (\mathbb{E}\{n^2\} - \mathbb{E}^2\{n\}) \tag{A31a}$$

$$= m(\frac{\tau}{2\pi})^2 \times V_H(\psi) , \qquad (A31b)$$

where $V_H(\psi) = \langle \psi | H^2 | \psi \rangle - \langle \psi | H | \psi \rangle^2$ is the energy variance of state ψ for Hamiltonian H. In particular, corollary 11 implies

c = c

$$d_{\text{TV}}(p_{\psi^{\otimes m}}, TP(\mu, \sigma^2)) \le \frac{c}{\sqrt{mb - 1/2}} + \frac{2}{m\sqrt{(\tau/2\pi)^2 \times V_H(\psi)}}$$
 (A32)

Here,

$$b = \min\{\frac{1}{2}, 1 - \frac{1}{2} \sum_{l} |p_{\psi}(l) - p_{\psi}(l+1)|\} , \qquad (A33)$$

where $\frac{1}{2}\sum_{l}|p_{\psi}(l)-p_{\psi}(l+1)|$ is the total variation distance between p_{ψ} and the translated version of p_{ψ} , and c, defined in corollary 11 is a finite number (independent of m), provided that the energy variance of ψ is nonzero and it has bounded third moment of energy.

It follows that the two distributions $p_{\psi^{\otimes m}}$ and $TP(\mu, \sigma^2)$ converge in the total variation distance, if

- 1. The distribution p_{ψ} has a nonzero variance, which means ψ is not an eigenstate of the system Hamiltonian H.
- 2. The distribution p_{ψ} has a finite third moment (This also guarantees that coefficient c in Eq.(A32) is finite).
- 3. The total variation distance between p_{ψ} and translated version of p_{ψ} satisfies

$$\frac{1}{2} \sum_{n} |p_{\psi}(n) - p_{\psi}(n+1)| < 1 , \qquad (A34)$$

which means the two distributions $p_{\psi}(n)$ and $\tilde{p}_{\psi}(n) = p_{\psi}(n+1)$ have overlapping supports. This condition is satisfied if there is, at least, an integer $n_0 \in \mathbb{Z}$ such that both $p_{\psi}(n_0)$ and $p_{\psi}(n_0+1)$ are non-zero.

Next, we use this result to study state conversion in the iid regime.

5. State Conversion in the iid regime: Special case

In this section we study state conversion in the iid regime for the special case of states which satisfy condition 3 in Eq.(A34) above. Later, in the next section we show how this constraint can be relaxed.

Consider two systems with states ψ_1 and ψ_2 and Hamiltonians H_1 and H_2 , respectively. Suppose both systems have period τ , i.e.

$$|\langle \psi_{1,2}|e^{-i\tau H_{1,2}}|\psi_{1,2}\rangle| = 1$$
. (A35)

Furthermore, assume the energy references are chosen such that

$$e^{-i\tau H_{1,2}}|\psi_{1,2}\rangle = |\psi_{1,2}\rangle$$
 (A36)

Assume conditions 1-3 above are satisfied for both states $\psi_{1,2}$. Let

$$R = \frac{V_{H_1}(\psi_1)}{V_{H_2}(\psi_2)},\tag{A37}$$

be the ratio of their energy variances. For any integer m suppose $p_{\psi_1^{\otimes m}}$ be the energy distribution for state $\psi_1^{\otimes m}$ and $p_{\psi_2^{\otimes \lceil Rm \rceil}}$ be the energy distribution for state $\psi_2^{\otimes \lceil Rm \rceil}$. Note that $p_{\psi_1^{\otimes m}}$ is equal to the m-fold convolution of p_{ψ_1} , i.e.

$$p_{\psi_1^{\otimes m}} = \underbrace{p_{\psi_1} * \dots * p_{\psi_1}}_{m \text{ times}}. \tag{A38}$$

Similarly,

$$p_{\psi_2^{\otimes \lceil Rm \rceil}} = \underbrace{p_{\psi_2} * \dots * p_{\psi_2}}_{\lceil Rm \rceil \text{ times}}.$$
 (A39)

From corollary 11 and the discussion in the previous section we know that, assuming conditions 1-3 are satisfied in the limit of large m the energy distribution for state $\psi_1^{\otimes m}$ converges to the translated Poisson distribution $TP(\mu_1, \sigma_1^2)$, where

$$\mu_1 = m(\frac{\tau}{2\pi}) \times \langle \psi_1 | H_1 | \psi_1 \rangle , \qquad (A40)$$

$$\sigma_1^2 = m(\frac{\tau}{2\pi})^2 \times V_{H_1}(\psi_1) = m(\frac{\tau}{2\pi})^2 \times \left[\langle \psi_1 | H_1^2 | \psi_1 \rangle - \langle \psi_1 | H_1 | \psi_1 \rangle^2 \right] . \tag{A41}$$

That is

$$\lim_{m \to \infty} d_{\text{TV}}(p_{\psi_1^{\otimes m}}, TP(\mu_1, \sigma_1^2)) = 0.$$
 (A42)

Similarly, in the limit of large m the energy distribution for state $\psi_2^{\otimes \lceil Rm \rceil}$ converges to the translated Poisson distribution $TP(\mu_2, \sigma_2^2)$, where

$$\mu_2 = m(\frac{\tau}{2\pi})\langle \psi_2 | H_2 | \psi_2 \rangle , \qquad (A43)$$

$$\sigma_2^2 = m(\frac{\tau}{2\pi})^2 \times V_{H_2}(\psi_2) = m(\frac{\tau}{2\pi})^2 \times \left[\langle \psi_2 | H_2^2 | \psi_2 \rangle - \langle \psi_2 | H_2 | \psi_2 \rangle^2 \right] . \tag{A44}$$

That is

$$\lim_{m \to \infty} d_{\text{TV}}(p_{\psi_2^{\otimes \lceil Rm \rceil}}, TP(\mu_2, \sigma_2^2)) = 0.$$
(A45)

Next, we show that in the limit of large m a translated version of the distribution $TP(\mu_1, \sigma_1^2)$ converges to the distribution $TP(\mu_2, \sigma_2^2)$ in the total variation distance. Recall that the distribution $TP(\mu, \sigma)$ is the distribution obtained from translating a Poisson distribution with variance $\sigma^2 + \gamma$, where $0 \le \gamma \le 1$. Therefore, up to a translation $TP(\mu_1, \sigma_1^2)$ and $TP(\mu_2, \sigma_2^2)$ are, respectively, equal to the Poisson distributions $P(\sigma_1^2 + \gamma_1)$ and $P(\sigma_2^2 + \gamma_2)$, where $0 \le \gamma_{1,2} \le 1$.

Next, we use lemma 12 which bounds the distance between two Poisson distributions, in terms of the difference between their variances. For the two distributions $P(\sigma_1^2 + \gamma_1)$ and $P(\sigma_2^2 + \gamma_2)$, the difference between the variances is

$$(\sigma_2^2 + \gamma_2) - (\sigma_1^2 + \gamma_1) = (\frac{\tau}{2\pi})^2 \times [\lceil Rm \rceil V_{H_2}(\psi_2) - mV_{H_1}(\psi_1)] + (\gamma_2 - \gamma_1)$$
(A46)

$$= \left(\frac{\tau}{2\pi}\right)^2 \times V_{H_1}(\psi_1) \left(\frac{\lceil Rm \rceil}{R} - m\right) + \left(\gamma_2 - \gamma_1\right). \tag{A47}$$

Using the facts that $|\gamma_1-\gamma_2|\leq 1$ and $|\frac{\lceil Rm\rceil}{R}-m|\leq 1/R$, we find

$$\lim_{m \to \infty} \frac{(\sigma_2^2 + \gamma_2) - (\sigma_1^2 + \gamma_1)}{\sigma_1} = \lim_{m \to \infty} \frac{(\frac{\tau}{2\pi})^2 V_{H_1}(\psi_1)(\frac{\lceil Rm \rceil}{R} - m) + (\gamma_2 - \gamma_1)}{\sqrt{m(\frac{\tau}{2\pi})^2 V_{H_1}(\psi_1)}} = 0.$$
 (A48)

Therefore, we conclude that if $R = \frac{V_H(\psi_1)}{V_H(\psi_2)} > 0$, then in the limit m goes to infinity, a properly translated version of the energy distribution of state $\psi_1^{\otimes m}$ converges to the energy distribution for state $\psi_2^{\otimes \lceil Rm \rceil}$ in the total variation distance. Combining this with proposition 9, we arrive at

Proposition 13. Consider two systems with Hamiltonian H_1 and H_2 and states ψ_1 and ψ_2 , respectively. Assume:

1. Both systems have period τ , such that

$$\tau = \inf_{t} \{ t > 0 : \left| \langle \psi_{1,2} | e^{-iH_{1,2}t} | \psi_{1,2} \rangle \right| = 1 \}. \tag{A49}$$

2. The energy distributions $p_{\psi_{1,2}}$ satisfy the condition

$$\frac{1}{2} \sum_{l} |p_{\psi_{1,2}}(l) - p_{\psi_{1,2}}(l+1)| < 1,$$
(A50)

where $p_{\psi_{1,2}}(n)=\frac{1}{2\pi}\int_0^{2\pi}d\theta~e^{i\theta n}\langle\psi_{1,2}|e^{-iH_{1,2}\tau\frac{\theta}{2\pi}}|\psi_{1,2}\rangle$, is the probability that state $\psi_{1,2}$ has energy $2\pi n/\tau$ with respect to Hamiltonian $H_{1,2}$, where we have defined the energy references for Hamiltonians H_1 and H_2 such that $e^{-iH_{1,2}\tau}|\psi_{1,2}\rangle=|\psi_{1,2}\rangle$.

Let $R = \frac{V_H(\psi_1)}{V_H(\psi_2)}$. Then, for any integer m there exists a TI transformation which maps $\psi_1^{\otimes m}$ to a state close to $\psi_2^{\otimes \lceil Rm \rceil}$, such that their trace distance vanishes in the limit m goes to infinity.

Next, we extend the result and show that condition in Eq.(A50) can be relaxed.

6. State conversion in the iid regime: General case

The above result was obtained under the assumption that the energy distribution p_{ψ} associated to state ψ satisfies condition

$$\frac{1}{2} \sum_{n} |p_{\psi}(n) - p_{\psi}(n+1)| < 1, \qquad (A51)$$

which means there is an integer n such that both $p_{\psi}(n)$ and $p_{\psi}(n+1)$ are nonzero. Now consider the general case where this condition is not satisfied, i.e. assume $p_{\psi}(n)$ and $\tilde{p}_{\psi}(n) = p_{\psi}(n+1)$ are perfectly distinguishable distributions.

This is the case, for example, for the energy distributions associated to states

$$|\eta\rangle = \frac{|0\rangle + |2\rangle}{\sqrt{2}}$$
 and $|\gamma\rangle = \frac{|0\rangle + |2\rangle + |5\rangle}{\sqrt{3}}$,

with the Hamiltonian $H=2\pi/\tau\sum_{k=0}^{\infty}k|k\rangle\langle k|$.

Nevertheless, one can easily see that although both the distributions p_{η} and p_{γ} do not satisfy Eq.(A51), there is an important distinction between them: Suppose instead of one copy of state $|\gamma\rangle$ we look at the energy distribution for two copies of this state, which is given by the distribution $p_{\gamma\otimes 2}=p_{\gamma}*p_{\gamma}$. This distribution has a nonzero support on n=0,2,4,5,7,10. It follows that, even though the energy distribution for one copy of γ does not satisfy Eq.(A51), energy distribution for two copies of this state *does* satisfy this condition. That is the total variation distance between $p_{\gamma}*p_{\gamma}(n)$ and its translated version $p_{\gamma}*p_{\gamma}(n+1)$ is less than one,

$$\frac{1}{2} \sum_{n} |p_{\gamma} * p_{\gamma}(n+1) - p_{\gamma} * p_{\gamma}(n)| < 1.$$
 (A52)

Thus, we can apply our result in the previous section to two copies of this state and conclude that, in the limit m goes to infinity, the energy distribution for $\gamma^{\otimes m}$ converges to a translated Poisson distribution.

On the other hand, this will not happen for state $|\eta\rangle$: Since the energy distribution of state $|\eta\rangle$ has support only on even integers n=0,2, it turns out that for any integer L, the energy distribution of state $\eta^{\otimes L}$ also has support only on even integers. Therefore, in the limit of large L, the energy distribution will not converge to a translated Poisson distribution (In fact, it converges to a translated Poisson distribution defined only on even integers).

It turns out that the distinction between these two examples have a simple physical interpretation, in terms of the period of dynamics. Recall that the period of dynamics for a system with state ψ and Hamiltonian H is defined as

$$\inf_{t}\{t>0: \left|\langle\psi|e^{-iHt}|\psi\rangle\right|=1\}\;. \tag{A53}$$

It can be easily seen that for state $|\eta\rangle$ the period of dynamics is $\tau/2$ and for state $|\gamma\rangle$ this period is τ . It follows that, in general, having the full period τ is the necessary and sufficient condition to guarantee that condition in Eq.(A51) is satisfied for a finite number of copies of state.

Lemma 14. Suppose under Hamiltonian H the period of state $|\psi\rangle$ is τ , i.e. $\tau=\inf_t\{t>0: |\langle\psi|e^{-iHt}|\psi\rangle|=1\}$. Let

$$p_{\psi}(n) = \frac{1}{2\pi} \int_{0}^{2\pi} d\theta \ e^{i\theta n} \langle \psi | e^{-iH\tau \frac{\theta}{2\pi}} | \psi \rangle \ , \tag{A54}$$

be the probability that state ψ has energy $2\pi n/\tau$, where we have chosen the energy reference such that $e^{-iH\tau}|\psi\rangle=|\psi\rangle$. Then, there is a finite L such that the distribution $p_{\psi\otimes L}=\underbrace{p_{\psi}*\cdots*p_{\psi}}_{L \text{ times}}$, corresponding to the energy distribution of $\psi^{\otimes L}$, satisfies

$$\frac{1}{2} \sum_{n \in \mathbb{Z}} |p_{\psi \otimes L}(n) - p_{\psi \otimes L}(n+1)| < 1.$$
(A55)

We prove this lemma at the end of this section, using Bezout's theorem.

Let L be an integer for which Eq.(A55) holds. Next, we consider the energy distribution for state $\psi^{\otimes mL}$ in the limit of large m, and apply the argument in Sec.(A4). According to this argument, if $\psi^{\otimes L}$ (1) has non-zero energy variance, (2) has finite third moment of energy, and (3) satisfies

$$\frac{1}{2} \sum_{n \in \mathbb{Z}} |p_{\psi \otimes L}(n) - p_{\psi \otimes L}(n+1)| < 1, \qquad (A56)$$

then in the limit of large m the energy distribution $p_{\psi^{\otimes mL}}$ converges to a translated Poisson distribution with variance $\tau^2 m L V_H(\psi)$. Note that $\psi^{\otimes L}$ has non-zero energy variance if ψ is not an eigenstate of energy, and $\psi^{\otimes L}$ has finite third moment of energy if the third moment is finite for ψ .

Then, combining this with the argument we used to prove proposition 13 in Sec.(A 5) we can prove the following theorem.

Theorem 15. Consider two systems with Hamiltonian H_1 and H_2 and states ψ_1 and ψ_2 , respectively. Assume both systems have period τ , i.e. $\tau = \inf_t \{t > 0 : \left| \langle \psi_{1,2} | e^{-iH_{1,2}t} | \psi_{1,2} \rangle \right| = 1 \}$. Let $R = \frac{V_H(\psi_1)}{V_H(\psi_2)}$. Then, for any integer m there exists a TI transformation which maps $\psi_1^{\otimes m}$ to a state close to $\psi_2^{\otimes \lceil Rm \rceil}$, such that their trace distance vanishes in the limit m goes to infinity.

Remark 16. For pure states, the energy variance is one fourth of QFI, i.e. $F_H(\psi) = 4V_H(\psi)$. In Section D, Theorem 20 shows that QFI is non-increasing in the asymptotic transformations with vanishing errors. In the case of above theorem this implies that the error vanishes in the asymptotic limit, if and only if the rate R satisfies $R \leq \frac{V_H(\psi_1)}{V_H(\psi_2)}$.

We finish this section by proving lemma 14.

Proof. Let $n_{\min}2\pi/\tau$ be the minimum occupied energy level by state ψ (Note that any Hamiltonian has a lowest energy level). In other words, let

$$n_{\min} = \min\{n : p_{\psi}(n) \neq 0\},$$
 (A57)

be the minimum n for which $p_{\psi}(n) \neq 0$. Let

$$\mathcal{N}_{\psi} = \{ n - n_{\min} : p_{\psi}(n) \neq 0 \}$$
 (A58)

be the set of all occupied levels shifted by n_{\min} . The fact that the period is τ implies that the greatest common divisor of this set is 1, i.e.

$$\gcd(\mathcal{N}_{\psi}) = 1. \tag{A59}$$

This can be seen by noting that if $k = \gcd(\mathcal{N}_{\psi})$, then for any n either $p_{\psi}(n) = 0$ or $n - n_{\min} = jk$ for an integer j. Therefore, since energy levels are related to integer n via relation $E = 2\pi n/\tau$, we find

$$|\langle \psi | e^{-iH\tau/k} | \psi \rangle| = |\sum_{n} p_{\psi}(n) e^{-i2\pi(jk + n_{\min})/k}| = |\sum_{n} p_{\psi}(n) e^{-i2\pi n_{\min}/k}| = 1,$$
(A60)

which implies the period of state τ/k .

Therefore, assuming the period is τ , we have $gcd(\mathcal{N}_{\psi}) = 1$.

Next, we use Bezout's theorem:

Lemma 17. (Bezout's theorem) Suppose the greatest common divisor of a set integers $\{a_1, \dots, a_n\}$ be one, i.e. $gcd(\{a_1, \dots, a_n\}) = 1$. Then, there exists integers $\{x_1, \dots, x_n\}$, such that $\sum_{i=1}^n x_i a_i = 1$.

We apply this result to the set of integers $\{n_i\}_i = \mathcal{N}_{\psi}$. Then, the fact that the greatest common divisor of this set is one implies that there exists a set of integers $\{x_i\}_i$ such that

$$\sum_{i} x_i n_i = 1. \tag{A61}$$

Partitioning the set $\{x_1, \dots, x_n\}$ to two subsets which only include positive and negative elements of this set, we find

$$\sum_{i:x_i>0} x_i n_i = 1 - \sum_{i:x_i<0} x_i n_i = 1 + \sum_{i:x_i<0} |x_i| n_i.$$
(A62)

Let $L = \sum_{i} |x_i|$ and consider the probability distribution

$$p_{\psi^{\otimes L}} = \underbrace{p_{\psi} * \cdots * p_{\psi}}_{L \text{ times}}$$

corresponding to the total energy distribution for state $\psi^{\otimes L}$. This is the probability distribution for the random variable $\sum_{r=1}^{L} N_r$, assuming each N_r has the distribution p_{ψ} . We show that for this distribution

$$\frac{1}{2} \sum_{n \in \mathbb{Z}} |p_{\psi \otimes L}(n) - p_{\psi \otimes L}(n+1)| < 1.$$
(A63)

Let

$$K = n_{\min}L + \sum_{i:x_i > 0} x_i n_i = n_{\min}L - \sum_{i:x_i < 0} x_i n_i + 1.$$
(A64)

Then, it can be easily seen that the random variable $\sum_{r=1}^L N_r$ takes both values K and K+1, with a non-zero probability. To see this, first consider the following event: For each $n_i \in \mathcal{N}_{\psi}$ with $x_i > 0$, suppose x_i different random variables in the set $\{N_r: 1 \leq r \leq L\}$ take the value $n_{\min} + n_i$, and the rest of the random variables, i.e. $L - \sum_{i:x_i>0} x_i$ random variables, take the value n_{\min} . In this event, the sum $\sum_{r=1}^L N_r$ will be equal to $K = n_{\min}L + \sum_{i:x_i>0} x_i n_i$. It follows that

$$p_{\eta_b \otimes L}(K) > 0. \tag{A65}$$

Next, consider a different event in which for each $x_i < 0$, $|x_i|$ different random variables in the set $\{N_r : 1 \le r \le L\}$ take the value $n_{\min} + n_i$, and the rest of random variables in this set, i.e. $L - \sum_{i:x_i < 0} |x_i|$, take the value n_{\min} . In this event the sum $\sum_{r=1}^{L} N_r$ will be equal to $K - 1 = n_{\min}L + \sum_{i:x_i < 0} |x_i| n_i$. It follows that

$$p_{\eta t \otimes L}(K-1) > 0. \tag{A66}$$

We conclude that the distribution $p_{\psi^{\otimes L}} = \underbrace{p_{\psi} * \cdots * p_{\psi}}_{L \text{ times}}$ is nonzero for both K and K-1. This immediately implies

$$\frac{1}{2} \sum_{n} |p_{\psi^{\otimes L}}(n) - p_{\psi^{\otimes L}}(n+1)| < 1 , \qquad (A67)$$

and proves the lemma. \Box

Appendix B: Quantum Fisher Information: Preliminaries

Here, we present a review of some useful properties of Quantum Fisher Information (QFI). See e.g. [8–10, 62, 63] for further details.

QFI for a general family of states ρ_t labeled by the real continuous parameter t is defined by

$$I_F(t) = \text{Tr}(\rho_t L_t^2) \,, \tag{B1}$$

where L_t is the symmetric logarithmic derivative, defined via equation

$$\dot{\rho}_t = \frac{1}{2} (\rho_t L_t + L_t \rho_t) . \tag{B2}$$

In the special case of $\rho_t = e^{-itH}\rho e^{itH}$ for a Hermitian operator H, we find

$$\dot{\rho}(t) = -i[H, \rho_t] = \frac{1}{2}(\rho_t L_t + L_t \rho_t)$$
 (B3)

Using the spectral decomposition of state ρ , as $\rho = \sum_k p_k |\phi_k\rangle \langle \phi_k|$ we find

$$2i \times \frac{p_k - p_j}{p_k + p_j} \langle \phi_k | L_t | \phi_j \rangle = \langle \phi_k | e^{iHt} L_t e^{-iHt} | \phi_j \rangle.$$
 (B4)

Putting this back into Eq.(B1) we find

$$I_F(t) = \text{Tr}(\rho_t L_t^2) = \text{Tr}(\rho L_0^2) = I_F,$$
 (B5)

i.e. the QFI is independent of the parameter t, and therefore we denote it by I_F . Then, it can be easily seen that

$$I_F = I_F(t) = \text{Tr}(\rho_t L_t^2) \tag{B6a}$$

$$= \sum_{k,j} p_k |\langle \phi_k | L_0 | \phi_j \rangle|^2 \tag{B6b}$$

$$=4\sum_{k,j}p_{k}\frac{(p_{k}-p_{j})^{2}}{(p_{k}+p_{j})^{2}}|\langle\phi_{k}|H|\phi_{j}\rangle|^{2}$$
(B6c)

$$=2\sum_{k,j}(p_k+p_j)\frac{(p_k-p_j)^2}{(p_k+p_j)^2}|\langle\phi_k|H|\phi_j\rangle|^2$$
(B6d)

$$= 2\sum_{k,j} \frac{(p_k - p_j)^2}{p_k + p_j} |\langle \phi_k | H | \phi_j \rangle|^2 .$$
 (B6e)

Note that if ρ is not full rank, we can apply the above formula to the state $\rho_{\epsilon} = (1 - \epsilon)\rho + \epsilon I/d$ for a vanishing $\epsilon \to 0$, where I/d is the totally mixed state. Using this technique, or applying the definition in Eq.(B3) we find that for pure states QFI is four time the variance of state ψ with the respect to the observable H, i.e.

$$I_F = 4 \times (\langle \psi | H^2 | \psi \rangle - \langle \psi | H | \psi \rangle^2) . \tag{B7}$$

In the following, we use the notation $F_H(\rho)$ to denote QFI for the family of state $e^{-iHt}\rho e^{iHt}$. In summary, for a system with state ρ and Hamiltonian H, QFI is given by

$$F_H(\rho) = 2\sum_{i,j} \frac{(p_k - p_j)^2}{p_k + p_j} |\langle \phi_k | H | \phi_j \rangle|^2 .$$
 (B8)

1. Review of some useful properties of Quantum Fisher Information

QFI is closely related to the fidelity. Let

$$Fid(\rho, \sigma) \equiv Tr(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}) = \|\sqrt{\rho}\sqrt{\sigma}\|_1,$$
(B9)

be the fidelity of states ρ and σ . Consider the fidelity of state ρ and $e^{-iHt}\rho e^{iHt}$ as a function of t. Then, it can be easily seen that at t=0, the first derivative of this function with respect to t vanishes, i.e.

$$\frac{d}{dt} \operatorname{Fid}(\rho, e^{-iHt} \rho e^{iHt}) \Big|_{t=0} = 0.$$
(B10)

Furthermore, it turns out that the second derivative is given by QFI, i.e.

$$F_H(\rho) = -4 \frac{d^2}{dt^2} \operatorname{Fid}(\rho, e^{-iHt} \rho e^{iHt}) \Big|_{t=0}. \tag{B11}$$

Therefore, roughly speaking, QFI determines how fast states ρ and $e^{-iHt}\rho e^{iHt}$ become distinguishable. QFI has the following important properties [8–10, 62, 63]:

1. Faithfulness: It is zero if, and only if, state is incoherent, i.e. diagonal in the energy eigenbasis. This can be seen using the fact that $[\rho, H] = 0$ if and only if for all i and j,

$$\langle \phi_i | [\rho, H] | \phi_i \rangle = (p_i - p_i) \langle \phi_i | H | \phi_i \rangle = 0. \tag{B12}$$

Using the formula

$$F_H(\rho) = 2\sum_{i,j} \frac{[(p_i - p_j)|\langle \phi_i | H | \phi_j \rangle|]^2}{p_i + p_j} , \qquad (B13)$$

we can easily see that this is the case if, and only if, $F_H(\rho) = 0$.

2. Monotonicity: It is non-increasing under any TI operation \mathcal{E}_{TI} , i.e.

$$F_H(\mathcal{E}_{TI}(\rho)) \le F_H(\rho).$$
 (B14)

In particular, it remains invariant under energy-conserving unitaries. This can be easily seen, e.g. using the connection between QFI and the fidelity, and the fact that fidelity satisfies information processing inequality, i.e.

$$\operatorname{Fid}(\rho, \sigma) \le \operatorname{Fid}(\mathcal{E}(\rho), \mathcal{E}(\sigma))$$
, (B15)

for any trace-preserving completely positive map \mathcal{E} .

- 3. Additivity: For a composite non-interacting system with the total Hamiltonian $H_{\text{tot}} = H_1 \otimes I_2 + I_1 \otimes H_2$, QFI is additive for uncorrelated states, i.e. $F_{H_{\text{tot}}}(\rho_1 \otimes \rho_2) = F_{H_1}(\rho_1) + F_{H_2}(\rho_2)$. This can be seen, e.g. from the multiplicativity of the fidelity for tensor products, together with the connection between fidelity and QFI in Eq.(B11).
- 4. Convexity: For any $0 \le p \le 1$ and states ρ and σ , $F_H(p\rho + (1-p)\sigma) \le pF_H(\rho) + (1-p)F_H(\sigma)$. This also can be seen from the concavity of the fidelity together with the connection between fidelity and QFI in Eq.(B11).

Appendix C: Quantum Fisher Information as the single-shot coherence cost

In this section we prove the following theorem.

Theorem 18. QFI of state ρ of system S is four times the minimum energy variance of all purifications of ρ with auxiliary systems not interacting with S, i.e.

$$F_H(\rho) = \min_{\Phi_{\rho}, H_A} F_{H_{tot}}(\Phi_{\rho}) = 4 \times \min_{\Phi_{\rho}, H_A} V_{H_{tot}}(\Phi_{\rho}), \qquad (C1)$$

where the minimization is over all pure states $|\Phi_{\rho}\rangle_{SA}$ satisfying $Tr_A(|\Phi_{\rho}\rangle\langle\Phi_{\rho}|_{SA})=\rho$, and all Hamiltonians of A.

Then, as a simple corollary of this theorem, we present a new proof of the following theorem, which is originally conjectured (and proved in a special case) by Toth and Petz [70] and is proven by Yu [71].

Theorem 19. (Yu-Toth-Petz [70, 71]) *QFI is four times the* convex roof *of the variance, i.e.*

$$F_H(\rho) = \min_{\{p_i, \phi_i\}} \sum_i p_i F_H(\phi_i) = 4 \times \min_{\{p_i, \phi_i\}} \sum_i p_i V_H(\phi_i) , \qquad (C2)$$

where the minimization is over the set of all ensembles of pure states $\{p_i, \phi_i\}$ satisfying $\sum_i p_i |\phi_i\rangle \langle \phi_i| = \rho$.

1. proof of theorem 18

Consider system S with Hamiltonian H_S and state ρ with the spectral decomposition $\rho = \sum_i p_i |\phi_i\rangle \langle \phi_i|$. Consider an auxiliary system A with Hamiltonian H_A . Let $|\Phi_\rho\rangle$ be a pure state of systems A and S which purifies state ρ_S , such that

$$\rho_S = \text{Tr}_A(|\Phi_\rho\rangle\langle\Phi_\rho|),\tag{C3}$$

where the partial trace is over system A.

Let H_{tot} be the total Hamiltonian of the system S and auxiliary system A, i.e.

$$H_{\text{tot}} = H_S \otimes I_A + I_S \otimes H_A . \tag{C4}$$

We are interested in finding the purification $|\Phi_{\rho}\rangle$ and Hamiltonian H_A for which the total energy variance

$$V_{H_{\text{tot}}}(|\Phi_{\rho}\rangle) = \langle \Phi_{\rho}|H_{\text{tot}}^{2}|\Phi_{\rho}\rangle - \langle \Phi_{\rho}|H_{\text{tot}}|\Phi_{\rho}\rangle^{2} \tag{C5}$$

is minimized. Since all purifications of ρ are equal up to a unitary on system A we can fix the purification to be

$$|\Phi_{\rho}\rangle = \sum_{i} \sqrt{p_{i}} |\phi_{i}\rangle |\phi_{i}\rangle = (\sqrt{\rho} \otimes I) \sum_{i} |\phi_{i}\rangle |\phi_{i}\rangle , \qquad (C6)$$

and only vary the Hamiltonian H_A . For this purification the reduced state on system A is also state ρ , i.e.

$$\operatorname{Tr}_{S}(|\Phi_{\rho}\rangle\langle\Phi_{\rho}|) = \rho. \tag{C7}$$

Next, note that by adding a proper multiple of the identity operator to H_A , we can always make the expectation value of the total energy zero, such that

$$\langle \Phi_{\rho} | H_{\text{tot}} | \Phi_{\rho} \rangle = 0. \tag{C8}$$

But, adding a multiple of the identity operator to the Hamiltonian does not change the energy variance. Therefore, in the following, without loss of generality, we assume the expectation value of the total Hamiltonian H_{tot} is zero. This means that the energy variance is given by the following expectation value

$$V_{H_{\text{tot}}}(|\Phi_{\rho}\rangle) = \langle \Phi_{\rho}|H_{\text{tot}}^{2}|\Phi_{\rho}\rangle \tag{C9}$$

$$= \langle \Phi_{o} | H_{S}^{2} \otimes I_{A} | \Phi_{o} \rangle + \langle \Phi_{o} | I_{S} \otimes H_{A}^{2} | \Phi_{o} \rangle + 2 \langle \Phi_{o} | H_{S} \otimes H_{A} | \Phi_{o} \rangle . \tag{C10}$$

Then, using $|\Phi_{\rho}\rangle = (\sqrt{\rho} \otimes I) \sum_{i} |\phi_{i}\rangle |\phi_{i}\rangle$ we find

$$V_{H_{\text{tot}}}(|\Phi_{\rho}\rangle) = \text{Tr}(\rho H_S^2) + \text{Tr}(\rho H_A^2) + 2\text{Tr}(\sqrt{\rho}H_S\sqrt{\rho}H_A^T)$$
 (C11a)

$$= \operatorname{Tr}(\rho H_S^2) + \operatorname{Tr}(\rho (H_A^T)^2) + 2\operatorname{Tr}(\sqrt{\rho} H_S \sqrt{\rho} H_A^T), \qquad (C11b)$$

where T denotes the transpose relative to the eigenbasis of ρ , i.e. $\{|\phi_j\rangle\}_j$. Here, to get the second line we have used $\mathrm{Tr}(\rho(H_A^T)^2)=\mathrm{Tr}(\rho H_A^2)$ which follows from the fact that the trace of any operator remains invariant under transpose, together with the fact that ρ is diagonal in $\{|\phi_j\rangle\}_j$ basis, and so $\rho^T=\rho$.

Next, we consider small variations of H_A^T , denoted by δH_A^T . At the point where the variance $V_{H_{\text{tot}}}(|\Phi_{\rho}\rangle)$ is minimized, we have

$$\frac{\delta V_{H_{\text{tot}}}(|\Phi_{\rho}\rangle)}{\delta H_{A}^{T}} = \frac{\delta \langle \Phi_{\rho}|H_{\text{tot}}^{2}|\Phi_{\rho}\rangle}{\delta H_{A}^{T}} = 0. \tag{C12}$$

It can be easily seen that

$$\delta V_{H_{\text{tot}}}(|\Phi_{\rho}\rangle) = \delta \langle \Phi_{\rho} | H_{\text{tot}}^2 | \Phi_{\rho} \rangle \tag{C13a}$$

$$= \left[\text{Tr}(\rho(\delta H_A^T) H_A^T) + \text{Tr}(\rho H_A^T \delta H_A^T) + 2 \text{Tr}(\sqrt{\rho} H_S \sqrt{\rho} \delta H_A^T) \right] \tag{C13b}$$

$$+\mathcal{O}((\delta H_A^T)^2).$$
 (C13c)

At the point where the variance is minimized, this variation vanishes up to the first order with respect to δH_A^T , for all variations δH_A^T . This leads to the equation

$$\frac{H_A^T \rho + \rho H_A^T}{2} = -\sqrt{\rho} H_S \sqrt{\rho} , \qquad (C14)$$

which should be satisfied by ${\cal H}_A^T$ for which the variance is minimized.

Next, we find H_A^T which satisfies this equation. To solve this equation we vectorize both side, using the relation

$$Y = \sum_{i,j} Y_{i,j} |\phi_i\rangle \langle \phi_j| \longleftrightarrow \text{vec}(Y) = \sum_{i,j} Y_{i,j} |\phi_i\rangle |\phi_j\rangle , \qquad (C15)$$

which implies

$$\operatorname{vec}(XYZ) = (X \otimes Z^T)\operatorname{vec}(Y). \tag{C16}$$

Using this notation we can rewrite the above equation as

$$[I \otimes \rho^T + \rho \otimes I] \operatorname{vec}(H_A^T) = -2[\sqrt{\rho} \otimes \sqrt{\rho^T}] \operatorname{vec}(H_S).$$
 (C17)

This equation implies

$$\operatorname{vec}(H_A^T) = -2[I \otimes \rho^T + \rho \otimes I]^{-1} [\sqrt{\rho} \otimes \sqrt{\rho}^T] \operatorname{vec}(H_S) . \tag{C18}$$

Using the decomposition $\rho = \sum_i p_i |\phi_i\rangle\langle\phi_i|$ we find

$$\operatorname{vec}(H_A^T) = -2[I \otimes \rho^T + \rho \otimes I]^{-1}[\sqrt{\rho} \otimes \sqrt{\rho^T}]\operatorname{vec}(H_S)$$
(C19)

$$= -2 \left[\sum_{i,j} (p_i + p_j) |\phi_i\rangle \langle \phi_i| \otimes |\phi_j\rangle \langle \phi_j|^T \right]^{-1} [\sqrt{\rho} \otimes \sqrt{\rho}^T] \text{vec}(H_S)$$
 (C20)

$$= -2 \left[\sum_{i,j} (p_i + p_j)^{-1} |\phi_i\rangle \langle \phi_i| \otimes |\phi_j\rangle \langle \phi_j|^T \right] [\sqrt{\rho} \otimes \sqrt{\rho}^T] \text{vec}(H_S)$$
 (C21)

$$= -2 \left[\sum_{i,j} \frac{\sqrt{p_i p_j}}{p_i + p_j} |\phi_i\rangle \langle \phi_i| \otimes |\phi_j\rangle \langle \phi_j|^T \right] \operatorname{vec}(H_S) . \tag{C22}$$

Using Eq.(C16) this implies

$$H_A^T = -2\sum_{i,j} \frac{\sqrt{p_i p_j}}{p_i + p_j} |\phi_i\rangle\langle\phi_i| H_S |\phi_j\rangle\langle\phi_j| , \qquad (C23)$$

or, equivalently,

$$H_A = -2\sum_{i,j} \frac{\sqrt{p_i p_j}}{p_i + p_j} |\phi_j\rangle\langle\phi_i| H_S |\phi_j\rangle\langle\phi_i|.$$
 (C24)

Note that

$$Tr(\rho H_A) = -2\sum_{i,j} \frac{\sqrt{p_i p_j}}{p_i + p_j} Tr(\rho |\phi_j\rangle \langle \phi_i | H_S |\phi_j\rangle \langle \phi_i |)$$
 (C25a)

$$= -\sum_{i} p_{i} \langle \phi_{i} | H_{S} | \phi_{i} \rangle \tag{C25b}$$

$$= -\text{Tr}(\rho H_S). \tag{C25c}$$

It follows that the expectation value of the total Hamiltonian is zero, i.e. $\langle \Phi_{\rho}|H_{\rm tot}|\Phi_{\rho}\rangle=0.$

For this optimal H_A we have

$$\operatorname{Tr}(\rho H_A^2) = 4\operatorname{Tr}(\rho \left[\sum_{i,j} \frac{\sqrt{p_i p_j}}{p_i + p_j} |\phi_j\rangle \langle \phi_i | H_S |\phi_j\rangle \langle \phi_i | \right] \left[\sum_{k,l} \frac{\sqrt{p_k p_l}}{p_k + p_l} |\phi_l\rangle \langle \phi_k | H_S |\phi_l\rangle \langle \phi_k | \right]) \tag{C26a}$$

$$=4\sum_{i,j} \frac{p_i p_j^2}{(p_i + p_j)^2} |\langle \phi_i | H_S | \phi_j \rangle|^2$$
 (C26b)

$$=2\sum_{i,j}\frac{p_{i}p_{j}^{2}+p_{j}p_{i}^{2}}{(p_{i}+p_{j})^{2}}|\langle\phi_{i}|H_{S}|\phi_{j}\rangle|^{2}$$
(C26c)

$$=2\sum_{i,j}\frac{p_ip_j}{p_i+p_j}|\langle\phi_i|H_S|\phi_j\rangle|^2,$$
(C26d)

where to get the third line we have used the fact that $\frac{1}{(p_i+p_j)^2}|\langle\phi_i|H_S|\phi_j\rangle|^2$ is symmetric with respect to i and j.

Similarly,

$$\operatorname{Tr}(\sqrt{\rho}H_{S}\sqrt{\rho}H_{A}^{T}) = -2\operatorname{Tr}\left(H_{S}\sqrt{\rho}\sum_{i,j}\frac{\sqrt{p_{i}p_{j}}}{p_{i}+p_{j}}|\phi_{i}\rangle\langle\phi_{i}|H_{S}|\phi_{j}\rangle\langle\phi_{j}|\sqrt{\rho}\right) \tag{C27a}$$

$$= -2\operatorname{Tr}\left(H_S \sum_{i,j} \frac{p_i p_j}{p_i + p_j} |\phi_i\rangle\langle\phi_i| H_S |\phi_j\rangle\langle\phi_j|\right) \tag{C27b}$$

$$=-2\sum_{i,j}\frac{p_ip_j}{p_i+p_j}|\langle\phi_i|H_S|\phi_j\rangle|^2 \tag{C27c}$$

$$= -\text{Tr}(\rho H_A^2) \;, \tag{C27d}$$

where to get the last line we have used Eq.(C26).

Putting these into Eq.(C11a) we find

$$V_{H_{\text{tot}}}(|\Phi_{\rho}\rangle) = \text{Tr}(\rho H_S^2) + \text{Tr}(\rho H_A^2) + 2\text{Tr}(\sqrt{\rho}H_S\sqrt{\rho}H_A^T)$$
 (C28a)

$$= \text{Tr}(\rho H_S^2) - 2\sum_{i,j} \frac{p_i p_j}{p_i + p_j} |\langle \phi_i | H_S | \phi_j \rangle|^2$$
 (C28b)

$$= \sum_{i} p_i \langle \phi_i | H_S^2 | \phi_i \rangle - 2 \sum_{i,j} \frac{p_i p_j}{p_i + p_j} |\langle \phi_i | H_S | \phi_j \rangle|^2$$
 (C28c)

$$= \sum_{i,j} p_i |\langle \phi_i | H_S | \phi_j \rangle|^2 - 2 \sum_{i,j} \frac{p_i p_j}{p_i + p_j} |\langle \phi_i | H_S | \phi_j \rangle|^2$$
 (C28d)

$$= \sum_{i,j} \frac{(p_i + p_j)^2}{2(p_i + p_j)} |\langle \phi_i | H_S | \phi_j \rangle|^2 - 2 \sum_{i,j} \frac{p_i p_j}{p_i + p_j} |\langle \phi_i | H_S | \phi_j \rangle|^2$$
 (C28e)

$$= \sum_{i,j} \frac{(p_i - p_j)^2}{2(p_i + p_j)} |\langle \phi_i | H_S | \phi_j \rangle|^2 , \qquad (C28f)$$

where to get the fourth line we have used the decomposition of the identity operator as $\sum_j |\phi_j\rangle\langle\phi_j|$, and to get the fifth line we have used the fact that $|\langle\phi_i|H_S|\phi_j\rangle|^2$ is symmetric with respect to i and j.

Comparing this with the formula for QFI

$$F_H(\rho) = 2\sum_{i,j} \frac{(p_k - p_j)^2}{p_k + p_j} |\langle \phi_k | H | \phi_j \rangle|^2 , \qquad (C29)$$

we find that

$$V_{H_{\text{tot}}}(|\Phi_{\rho}\rangle) = \frac{1}{4}F_{H}(\rho) . \tag{C30}$$

This completes the proof.

a. Fisher information of the purifying system

The above argument shows that if the Hamiltonian of the auxiliary system is

$$H_A = -2\sum_{i,j} \frac{\sqrt{p_i p_j}}{p_i + p_j} |\phi_j\rangle\langle\phi_i| H_S |\phi_j\rangle\langle\phi_i|, \tag{C31}$$

then for the total Hamiltonian $H_S \otimes I_A + I_S \otimes H_A$ of the composite system S and A, the QFI of state $|\Phi_\rho\rangle = \sum_i \sqrt{p_i} |\phi_i\rangle |\phi_i\rangle$, is equal to the QFI for system S. In other words, by discarding system A the QFI does not decrease. It is interesting to note that this happens even though the QFI of the auxiliary system A is nonzero.

To calculate the QFI of the auxiliary system, first note that the reduced state of system A in this case is also ρ . Then, using the formula for QFI we find

$$F_{H_A}(\rho) = 2\sum_{i,j} \frac{(p_i - p_j)^2}{(p_i + p_j)} |\langle \phi_i | H_A | \phi_j \rangle|^2$$
 (C32)

$$= \sum_{i,j} \frac{2(p_i - p_j)^2}{(p_i + p_j)^2} \frac{4p_i p_j}{(p_i + p_j)^2} |\langle \phi_i | H_S | \phi_j \rangle|^2$$
 (C33)

$$= \sum_{i,j} \frac{8p_i p_j (p_i - p_j)^2}{(p_i + p_j)^3} |\langle \phi_i | H_S | \phi_j \rangle|^2$$
 (C34)

Therefore, if the system S is in a a full rank density operator with nonzero Fisher information, then the Fisher information for the auxiliary system will be necessarily nonzero, $F_{H_A}(\rho) > 0$.

We conclude that for state $|\Phi_{\rho}\rangle = \sum_{i} \sqrt{p_{i}} |\phi_{i}\rangle |\phi_{i}\rangle$, and for this choice of Hamiltonian H_{A} , by discarding system A, the Fisher information does not decrease, even though the process is irreversible, and the discarded system itself carries non-zero

Fisher information.

b. Comparison with the Wigner-Yanase skew Information

In the above argument we found the optimal Hamiltonian of auxiliary system for the joint state $|\Phi_{\rho}\rangle = \sum_i \sqrt{p_i} |\phi_i\rangle |\phi_i\rangle$. Since for this joint state the reduced state of both subsystems A and S is ρ , a natural choice for the Hamiltonian H_A which minimizes the total energy variance could be $H_A = -H_S^T$, where T denotes the transpose relative to the eigenbasis of ρ . Then, the total energy variance is given by

$$V_{H_{\text{tot}}}(|\Phi_{\rho}\rangle) = \langle \Phi_{\rho}|H_{\text{tot}}^2|\Phi_{\rho}\rangle - \langle \Phi_{\rho}|H_{\text{tot}}|\Phi_{\rho}\rangle^2 \tag{C35}$$

$$= \langle \Phi_{\rho} | H_S^2 \otimes I_A | \Phi_{\rho} \rangle + \langle \Phi_{\rho} | I_S \otimes H_A^2 | \Phi_{\rho} \rangle + 2 \langle \Phi_{\rho} | H_S \otimes H_A | \Phi_{\rho} \rangle \tag{C36}$$

$$= \operatorname{Tr}(\rho H_S^2) + \operatorname{Tr}(\rho H_A^2) + 2\operatorname{Tr}(\sqrt{\rho}H_S\sqrt{\rho}H_A^T)$$
(C37)

$$= 2\operatorname{Tr}(\rho H_S^2) - 2\operatorname{Tr}(\sqrt{\rho}H_S\sqrt{\rho}H_S), \tag{C38}$$

where, in the first line we have used the fact that for $H_A = -H_S^T$ the expectation value of total energy is zero. Interestingly, the last is twice the Wigner-Yanase skew information

$$W_H(\rho) = -\frac{1}{2} \text{Tr}\left([H_S, \sqrt{\rho}]^2\right), \tag{C39}$$

which is also a measure of asymmetry relative to time-translations. Therefore, for this choice of H_A we find

$$V_{H_{\text{tot}}}(|\Phi_{\rho}\rangle) = 2W_H(\rho). \tag{C40}$$

2. Theorem 19 as a corollary of theorem 18

Next, we prove theorem 19 using theorem 18 together with monotonicity of QFI under TI operations.

For a given system S with state ρ and Hamiltonian H_S , let $|\Phi\rangle_{SA}$ be the optimal purification of ρ obtained in theorem 18, which satisfies

$$F_{H_S}(\rho) = F_{H_{\text{tot}}}(\Phi) = 4(\langle \Phi | H_{\text{tot}}^2 | \Phi \rangle_{SA} - \langle \Phi | H_{\text{tot}} | \Phi \rangle_{SA}^2), \qquad (C41)$$

where $H_{\text{tot}} = I_A \otimes H_S + H_A \otimes I_S$ is the total Hamiltonian and H_A is the Hamiltonian of the auxiliary system.

Now suppose we measure system A in the eigenbasis of its Hamiltonian H_A , and obtain eigenstate $|E_k\rangle$ with probability p_k . After such measurement, the joint state of systems SA is

$$\sigma_{SA} = \sum_{k} |E_k\rangle \langle E_k|_A (|\Phi\rangle \langle \Phi|_{SA}) |E_k\rangle \langle E_k|_A \tag{C42}$$

$$= \sum_{k} p_{k} |E_{k}\rangle \langle E_{k}|_{A} \otimes |\eta_{k}\rangle \langle \eta_{k}|_{S} , \qquad (C43)$$

where $|\eta_k\rangle_S = \langle E_k|\Phi\rangle_{SA}$ is the state of system S given that system A is projected to state $|E_k\rangle$.

Since the measurement in the energy eigenbasis is covariant under time translations, i.e. the map $\mathcal{E}(\cdot) = \sum_k |E_k\rangle\langle E_k|_A(\cdot)|E_k\rangle\langle E_k|_A$ satisfies the covariance condition $\mathcal{E}(e^{-iH_At}(\cdot)e^{iH_At}) = e^{-iH_At}\mathcal{E}(\cdot)e^{iH_At}$, then it cannot increase QFI, i.e.

$$F_{H_{\text{tot}}}(\sigma_{SA}) \le F_{H_{\text{tot}}}(|\Phi\rangle\langle\Phi|_{SA}) = F_{H_S}(\rho)$$
 (C44)

On the other hand, in state σ_{SA} by tracing over system A we obtain system S in state ρ , i.e.

$$\operatorname{Tr}_{A}(\sigma_{SA}) = \sum_{k} p_{k} |\eta_{k}\rangle\langle\eta_{k}|_{S} = \rho$$
 (C45)

Since QFI is non-increasing under partial trace, then $F_{H_S}(\rho) \leq F_{H_{tot}}(\sigma_{SA})$. Therefore, we conclude that

$$F_{H_{\text{tot}}}(\sigma_{SA}) = F_{H_S}(\rho) . \tag{C46}$$

But, given that orthogonal states $|E_k\rangle$ are eigenstates of Hamiltonian H_A , one can easily see that QFI for state σ_{SA} is the average of QFI for states $|\eta_k\rangle_S$, i.e.

$$F_{H_{\text{tot}}}(\sigma_{SA}) = \sum_{k} p_k F_{H_S}(\eta_k) . \tag{C47}$$

Therefore, we find that

$$F_{H_S}(\rho) = \sum_k p_k F_{H_S}(\eta_k) . \tag{C48}$$

Finally, we note that the convexity of QFI implies that for any ensemble $\{q_i,|\psi_i\rangle\}$ which satisfies $\sum_i q_i |\psi_i\rangle \langle \psi_i| = \rho$ it holds that $F_{H_S}(\rho) \leq \sum_i q_i F_{H_S}(\psi_i)$. Therefore, we conclude that

$$F_H(\rho) = \min_{\{q_i, \psi_i\}} \sum_i q_i F_{H_S}(\psi_i) = \sum_i p_i F_{H_S}(\eta_i) , \qquad (C49)$$

which completes the proof.

Appendix D: Monotonicity of Fisher information in approximate asymptotic transformations

Recall that for a family of states $\{e^{-iHt}\rho e^{iHt}\}_t$ corresponding to the time-translated versions of state ρ , Quantum Fisher Information is defined by

$$F_H(\rho) = 2\sum_{j,k} \frac{(p_j - p_k)^2}{p_j + p_k} |\langle \psi_j | H | \psi_k \rangle|^2 ,$$
 (D1)

where $\rho = \sum_{j} p_{j} |\psi_{j}\rangle \langle \psi_{j}|$ is the spectral decomposition of ρ .

Let \mathcal{H}_{in} and \mathcal{H}_{out} be the Hilbert spaces of one copy of the input and output systems, respectively and H_{in} and H_{out} be arbitrary Hermitian operators, e.g. Hamiltonians, on \mathcal{H}_{in} and \mathcal{H}_{out} . Define the superoperators $\mathcal{U}_{in}(t)$ and $\mathcal{U}_{out}(t)$ to be the translations generated by H_{in} and H_{out} , i.e.

$$\mathcal{U}_{\rm in}(t)[\rho] = e^{-iH_{\rm in}t}\rho e^{-iH_{\rm in}t} \tag{D2}$$

and

$$\mathcal{U}_{\text{out}}(t)[\rho] = e^{-iH_{\text{out}}t}\rho e^{-iH_{\text{out}}t} . \tag{D3}$$

Theorem 20. Suppose for any integer n there exists a CPTP transformation \mathcal{E}_n which transforms n copies of the input system to $\lceil Rn \rceil$ copies of the output system, for R > 0. More precisely, \mathcal{E}_n maps density operators on $\mathcal{H}_{in}^{\otimes n}$ to density operators on $\mathcal{H}_{out}^{\otimes \lceil Rn \rceil}$. Suppose the CPTP maps \mathcal{E}_n satisfies the following covariance condition

$$\forall t: \ \mathcal{U}_{out}^{\otimes \lceil Rn \rceil}(t) \circ \mathcal{E}_n = \mathcal{E}_n \circ \mathcal{U}_{in}^{\otimes n}(t) . \tag{D4}$$

Suppose \mathcal{E}_n maps state $\rho^{\otimes n}$ to $\sigma^{\otimes \lceil Rn \rceil}$ with a small error in the trace distance, which vanishes in the limit n goes to infinity, such that

$$\lim_{n \to \infty} \|\mathcal{E}_n(\rho^{\otimes n}) - \sigma^{\otimes \lceil Rn \rceil}\|_1 = 0.$$
 (D5)

Then, the rate R is upper-bounded by the the ratio of Fisher information for ρ and σ , i.e.

$$R \le \frac{F_{H_{in}}(\rho)}{F_{H_{out}}(\sigma)} \,. \tag{D6}$$

Before presenting the proof, we recall the Fuchs-van de Graaf inequality [31, 32, 105]: For any pair of density operators ρ_1 and ρ_2 it holds that

$$1 - \operatorname{Fid}(\rho_1, \rho_2) \le \frac{1}{2} \|\rho_1 - \rho_2\| \le \sqrt{1 - \operatorname{Fid}^2(\rho_1, \rho_2)} . \tag{D7}$$

This implies that the convergence in the trace distance, is equivalent to convergence in the fidelity, i.e. Eq.(D5) can be rewritten as

$$\lim_{n \to \infty} \operatorname{Fid}(\mathcal{E}_n(\rho^{\otimes n}), \sigma^{\otimes \lceil Rn \rceil}) = 1.$$
(D8)

1. Proof

To simplify the notation we assume the input and output system Hamiltonians are identical and they are both denoted by H. Generalization to the case where these systems are different is straightforward.

For any n let $m = \lceil Rn \rceil$ be the number of copies of the output systems. Let $\sigma_m = \mathcal{E}_n(\rho^{\otimes n})$ be the actual output state and $\sigma_m(\Delta t)$ be the time-translated version of σ_m , i.e.

$$\sigma_m(\Delta t) = \mathcal{U}^{\otimes m}(\Delta t)[\sigma_m] \tag{D9}$$

$$= (e^{-iH\Delta t})^{\otimes m} \sigma_m (e^{iH\Delta t})^{\otimes m} . \tag{D10}$$

Here, Δt is a parameter whose value will be fixed later. Similarly, let $\sigma^{\otimes m}(\Delta t) = (e^{-iH\Delta t}\sigma e^{iH\Delta t})^{\otimes m}$ be the time-evolved

version of state $\sigma^{\otimes m}$. Since the transformation \mathcal{E}_n is covariant, i.e. satisfies Eq.(D4), for any Δt it maps state $\rho^{\otimes n}(\Delta t)$ to state $\sigma_m(\Delta t)$. To summarize

$$\mathcal{E}_n(\rho^{\otimes n}) = \sigma_m \tag{D11a}$$

$$\mathcal{E}_n(\rho(\Delta t)^{\otimes n}) = \sigma_m(\Delta t) . \tag{D11b}$$

Then, using the monotonicity of Fidelity under CPTP maps we find

$$\operatorname{Fid}(\sigma_m, \sigma_m(\Delta t)) \ge \operatorname{Fid}(\rho^{\otimes n}, \rho(\Delta t)^{\otimes n}) = \operatorname{Fid}^n(\rho, \rho(\Delta t)), \tag{D12}$$

where the equality follows from the multiplicativity of fidelity under tensor products.

Next, consider the Taylor expansion of the fidelity $\operatorname{Fid}(\rho, \rho(\Delta t))$ for small Δt . It can be easily seen that this is an even function of Δt , and therefore all of its odd derivatives with respect to Δt vanishes. Furthermore, it is known that the second derivative of function

$$\operatorname{Fid}(e^{-iH\Delta t}\rho e^{iH\Delta t},\rho) = \operatorname{Tr}\left(\sqrt{\sqrt{\rho}e^{-iH\Delta t}\rho e^{iH\Delta t}\sqrt{\rho}}\right) \tag{D13}$$

with respect to Δt is 1/4 times the Fisher information for the family of states $\{e^{-iH\Delta t}\rho e^{iH\Delta t}\}$ and parameter Δt , which we denote it by $F_H(\rho)$ (Theorem 6.3 [64]). It follows that

$$\operatorname{Fid}(e^{-iH\Delta t}\rho e^{iH\Delta t}, \rho) = 1 - \frac{\Delta t^2}{8}F_H(\rho) + \mathcal{O}(\Delta t^4), \qquad (D14)$$

where $\mathcal{O}(\Delta t^4)$ denotes terms of order Δt^4 and higher.

Next, let t be a constant real number, independent of n, and let $\Delta t = t/\sqrt{n}$. This implies

$$\operatorname{Fid}(e^{-iH\Delta t}\rho e^{iH\Delta t}, \rho) = 1 - \frac{\Delta t^2}{8}F_H(\rho) + \mathcal{O}(\Delta t^4)$$
(D15)

$$=1 - \frac{t^2}{8n} F_H(\rho) + \mathcal{O}(\frac{t^4}{n^2}).$$
 (D16)

Then, using the fact that $\lim_{n\to\infty} (1-x/n)^n = e^{-x}$, we find that in the limit n goes to infinity, $\operatorname{Fid}^n(\rho, \rho(\Delta t))$ in the right-hand side of Eq.(D12) converges to

$$\lim_{n \to \infty} \operatorname{Fid}^{n}(\rho, \rho(\frac{t}{\sqrt{n}})) = e^{-t^{2} F_{H}(\rho)/8} . \tag{D17}$$

Next, we focus on the left-hand side of Eq.(D12) and find a lower bound on $\operatorname{Fid}(\sigma_m, \sigma_m(\Delta t))$. Using the properties of fidelity and Bures metric we later prove the following lemma:

Lemma 21. For any pairs of states τ_1 and τ_2 and unitary U it holds that

$$\left| \operatorname{Fid}(U\tau_1 U^{\dagger}, \tau_1) - \operatorname{Fid}(U\tau_2 U^{\dagger}, \tau_2) \right| \le 4\sqrt{1 - \operatorname{Fid}(\tau_1, \tau_2)} . \tag{D18a}$$

Applying this result to states $\sigma^{\otimes m}$ and σ_m , we find

$$\left| \operatorname{Fid} \left(\sigma(\Delta t)^{\otimes m}, \sigma^{\otimes m} \right) - \operatorname{Fid} \left(\sigma_m(\Delta t), \sigma_m \right) \right| \le 4\sqrt{1 - \operatorname{Fid}(\sigma_m, \sigma^{\otimes m})} . \tag{D19}$$

Using the multiplicativity of fidelity for tensor products, this implies

$$\left| \operatorname{Fid}^{m}(\sigma(\Delta t), \sigma) - \operatorname{Fid}(\sigma_{m}(\Delta t), \sigma_{m}) \right| \leq 4\sqrt{1 - \operatorname{Fid}(\sigma_{m}, \sigma^{\otimes m})} . \tag{D20}$$

Assuming $\Delta t = t/\sqrt{n}$, and using the fact that $m = \lceil Rn \rceil$, we find that

$$\left| \operatorname{Fid}^{\lceil Rn \rceil} (\sigma(t/\sqrt{n}), \sigma) - \operatorname{Fid}(\sigma_m(t/\sqrt{n}), \sigma_m) \right| \le 4\sqrt{1 - \operatorname{Fid}(\sigma_m, \sigma^{\otimes m})} . \tag{D21}$$

Then, taking the limit $n \to \infty$ and using the fact that $\lim_{n \to \infty} \operatorname{Fid}(\sigma_m, \sigma^{\otimes m}) = 1$, we find that

$$\lim_{n \to \infty} \operatorname{Fid}^{\lceil Rn \rceil}(\sigma(t/\sqrt{n}), \sigma) = \lim_{n \to \infty} \operatorname{Fid}(\sigma_m(t/\sqrt{n}), \sigma_m). \tag{D22}$$

Using Eq.(D17) we find that the left-hand side converges to

$$\lim_{n \to \infty} \operatorname{Fid}^{\lceil Rn \rceil}(\sigma(t/\sqrt{n}), \sigma) = e^{-t^2 R F_H(\sigma)/8} . \tag{D23}$$

Therefore, the right-hand side of Eq.(D22) also converges to

$$\lim_{n \to \infty} \operatorname{Fid}(\sigma_m(t/\sqrt{n}), \sigma_m) = \lim_{n \to \infty} \operatorname{Fid}^{\lceil Rn \rceil}(\sigma(t/\sqrt{n}), \sigma) = e^{-t^2 R F_H(\sigma)/8} . \tag{D24}$$

Combining this with Eq.(D12) and Eq.(D17) we find

$$e^{-t^2 F_H(\rho)/8} \le e^{-t^2 R F_H(\sigma)/8},$$
 (D25)

for all $t \in \mathbb{R}$, which implies

$$F_H(\rho) \ge RF_H(\sigma),$$
 (D26)

and completes the proof.

2. Proof of lemma 21

We first recall some useful properties of Fidelity and the Bures metric. Recall that fidelity of two states ρ_1 and ρ_2 is defined as $\text{Fid}(\rho_1, \rho_2) = \|\sqrt{\rho_1}\sqrt{\rho_2}\|_1 = \text{Tr}(\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}})$. Fidelity is not a metric but it is closely related to the Bures metric, via the relation

$$B(\rho_1, \rho_2) = \sqrt{2}\sqrt{1 - \text{Fid}(\rho_1, \rho_2)}.$$
 (D27)

This immediately implies that Bures metric is invariant under exchange of states, i.e. $B(\rho_1, \rho_2) = B(\rho_2, \rho_1)$, and is also invariant under unitary transformations, i.e. $B(\rho_1, \rho_2) = B(U\rho_1 U^{\dagger}, U\rho_2 U^{\dagger})$. Furthermore, as a metric, it satisfies the triangle inequality, i.e.

$$B(\rho_1, \rho_2) + B(\rho_2, \rho_3) > B(\rho_1, \rho_3)$$
 (D28)

Now, we are ready to present the proof of the lemma.

Using the triangle inequality twice we find

$$B(U\tau_2U^{\dagger}, U\tau_1U^{\dagger}) + B(U\tau_1U^{\dagger}, \tau_1) + B(\tau_1, \tau_2) \ge B(U\tau_2U^{\dagger}, \tau_2).$$
 (D29)

Let $\eta \equiv B(U\tau_1U^{\dagger}, U\tau_2U^{\dagger}) = B(\tau_1, \tau_2)$. Then, the above inequality can be rewritten as $B(U\tau_1U^{\dagger}, \tau_1) \geq B(U\tau_2U^{\dagger}, \tau_2) - 2\eta$, which in turn implies

$$B^{2}(U\tau_{1}U^{\dagger}, \tau_{1}) \ge B^{2}(U\tau_{2}U^{\dagger}, \tau_{2}) - 4\eta B(U\tau_{2}U^{\dagger}, \tau_{2})$$
 (D30)

$$> B^2(U\tau_2U^{\dagger}, \tau_2) - 4\sqrt{2}\eta$$
 (D31)

where we have used the fact that Bures metric is bounded by $\sqrt{2}$. This, in turn implies

$$1 - \operatorname{Fid}(U\tau_1 U^{\dagger}, \tau_1) \ge 1 - \operatorname{Fid}(U\tau_2 U^{\dagger}, \tau_2) - 2\sqrt{2}\eta, \tag{D32}$$

and so

$$2\sqrt{2}\eta \ge \operatorname{Fid}(U\tau_1 U^{\dagger}, \tau_1) - \operatorname{Fid}(U\tau_2 U^{\dagger}, \tau_2). \tag{D33}$$

Exchanging τ_1 and τ_2 we also find $2\sqrt{2}\eta \geq \operatorname{Fid}(U\tau_2U^{\dagger},\tau_2) - \operatorname{Fid}(U\tau_1U^{\dagger},\tau_1)$, which completes the proof.

Appendix E: Quantum Fisher Information as the coherence cost: iid regime

In this section we prove that in the iid regime the coherence cost of preparing any state is determined by its Quantum Fisher information.

Theorem 22. Consider a system with Hamiltonian H and state ρ with period τ , with a finite-dimensional Hilbert space. Consider a two-level system with state $|\Phi\rangle_{c\text{-bit}} = (|0\rangle + |1\rangle)/\sqrt{2}$ and Hamiltonian $H_{c\text{-bit}} = \pi\sigma_z/\tau$. Then, for any $R > F_H(\rho)/F_{c\text{-bit}} = (\tau/2\pi)^2 F_H(\rho)$, and integer n, there exists a TI operation which transforms $\Phi_{c\text{-bit}}^{\otimes \lceil Rn \rceil}$ to a state whose trace distance from $\rho^{\otimes n}$ is bounded by $\epsilon > 0$, and $\epsilon \to 0$ in the limit n goes to infinity, i.e.

$$\Phi_{c\text{-}bit}^{\otimes \lceil Rn \rceil} \xrightarrow{TI} \stackrel{\epsilon}{\approx} \rho^{\otimes n} \text{ as } n \to \infty, \epsilon \to 0.$$

Furthermore, for any $R < F_H(\rho)/F_{c-bit} = (\tau/2\pi)^2 F_H(\rho)$ the above transformation is not possible with a vanishing error ϵ .

Another way to phrase this result is in terms of the coherence cost of state ρ .

Definition 23. The coherence cost (w.r.t. to TI operations) of a system with state ρ and Hamiltonian H is defined as

$$C_c^{T\!I}(\rho) = \inf R : \Phi_{c\text{-bit}}^{\otimes \lceil Rn \rceil} \xrightarrow{TI} \stackrel{\epsilon}{\approx} \rho^{\otimes n} \text{ as } n \to \infty, \epsilon \to 0 \ .$$

Then, the theorem implies that the coherence cost of state ρ is given by

$$C_c^{\text{TI}}(\rho) = F_H(\rho)/F_{\text{c-bit}}$$
 (E1)

The proof of the second part, i.e. $C_c^{\rm TI}(\rho) \geq F_H(\rho)/F_{\text{c-bit}}$ follows from our result in Sec.(D), and in particular, theorem 20 which shows Quantum Fisher Information cannot increase in the iid regime. In this section we prove the first part of this theorem, i.e. $C_c^{\rm TI}(\rho) \leq F_H(\rho)/F_{\text{c-bit}}$. The proof uses theorem 19, which proved in Sec. C (This theorem was originally conjectured and proved in a special case by Toth and Petz [70] and is proven by Yu [71]). According to this theorem QFI is four times the *convex roof* of the variance, i.e.

$$F_H(\rho) = \min_{\{p_i, \phi_i\}} \sum_i p_i F_H(\phi_i) = 4 \times \min_{\{p_i, \phi_i\}} \sum_i p_i V_H(\phi_i) , \qquad (E2)$$

where the minimization is over the set of all ensembles of pure states $\{p_i, \phi_i\}$ satisfying $\sum_i p_i |\phi_i\rangle \langle \phi_i| = \rho$.

In analogy with entanglement theory, the right-hand side of Eq.(E2) can be called the *coherence of formation* (w.r.t. TI operations) [72].

Using this result, together with the additivity of QFI, we can prove that QFI is an upper bound on coherence cost, i.e. $C_c^{\text{TI}}(\rho) \leq F_H(\rho)/F_{\text{c-bit}}$. The proof is very similar to the proof of similar results on entanglement cost [106] and coherence cost with respect to incoherence operations [6]: Briefly, to prepare state $\rho^{\otimes m}$ one can prepare the pure state $|\phi_{i_1}\rangle\cdots|\phi_{i_m}\rangle$ with probability $p_{i_1}\cdots p_{i_m}$, for a typical string $i_1\cdots i_m$, where $\rho=\sum_i p_i|\phi_i\rangle\langle\phi_i|$ is the optimal decomposition in Eq.(E2). We argue that the optimal decomposition can be chosen such that state $\bigotimes_i |\phi_i\rangle$ has period τ . Then, using an extension of theorem 19 it turns out that this state can be prepared by consuming, approximately, $\sum_i np_i 4V_H(\phi_i)/F_{\text{c-bit}}$ copies of the standard state $\Phi_{\text{c-bit}}$.

In the following, we present the details of the proof. First, we prove the following corollary of theorem 19.

Corollary 24. Suppose under Hamiltonian H state ρ has period τ , i.e. $\tau = \inf_t \{t > 0 : e^{-iHt} \rho e^{iHt} = \rho\}$. Then, there exists an ensemble $\{p_i, |\phi_i\rangle\}$ such that

(i)
$$\sum_{i} p_{i} |\phi_{i}\rangle\langle\phi_{i}| = \rho$$
, and

$$F_H(\rho) = \sum_i p_i F_H(\phi_i) = 4 \times \sum_i p_i V_H(\phi_i)$$
 (E3)

(ii) Let τ_i be the period of state $|\phi_i\rangle$, such that $\tau_i = \inf_t \{t > 0 : |\langle \phi_i|e^{-iHt}|\phi_i\rangle| = 1\}$. Then, either $\tau_i = 0$, i.e. state is an eigenstate of Hamiltonian H, or $\tau_i = \tau/k_i$ for an integer $k_i \in \mathbb{N}$. (iii) For any $t < \tau$, there is at least one state $|\phi_i\rangle$ such that $|\langle \phi_i|e^{-iHt}|\phi_i\rangle| \neq 1$. In other words, the greatest common divisor of

(iii) For any $t < \tau$, there is at least one state $|\phi_i\rangle$ such that $|\langle \phi_i|e^{-iHt}|\phi_i\rangle| \neq 1$. In other words, the greatest common divisor of integers $\{k_i = \tau/\tau_i\}$ is one.

Proof. First, note that $e^{-iH\tau}\rho e^{iH\tau}=\rho$ implies that for any two energy levels E_1 and E_2 if $\Pi_{E_1}\rho\Pi_{E_2}\neq 0$, then $E_1-E_2=2\pi k/\tau$ for an integer k. Based on the criterion that $\Pi_{E_1}\rho\Pi_{E_2}$ is zero or not we can divide the energy levels into disjoint partitions, such that (i) the energy levels in each partition are separated with energy gaps $2\pi k/\tau$ for an integer k, and (ii) for any two energy

levels E_1 and E_2 in two different partitions $\Pi_{E_1}\rho\Pi_{E_2}=0$. Suppose we label different partitions with label r, and let Π_r be the projector to the subspace spanned by the energy level in the sector r, i.e. each Π_r is the sum of Π_E for all E belonging to the same partition. Note that there is no coherence between different partitions, i.e.

$$\sum_{r} \tilde{\Pi}_{r} \rho \tilde{\Pi}_{r} = \rho . \tag{E4}$$

Let $\{p_i, |\phi_i\rangle\}$ be an optimal ensemble for which

$$F_H(\rho) = \min_{\{p_i, \phi_i\}} \sum_i p_i F_H(\phi_i) = 4 \times \min_{\{p_i, \phi_i\}} \sum_i p_i V_H(\phi_i) , \qquad (E5)$$

is satisfied. Now we define a new ensemble of pure states, which is obtained from this ensemble by removing coherence between different partitions defined above. Consider the ensemble

$$\left\{ \tilde{p}_{i,r} = p_i \langle \phi_i | \tilde{\Pi}_r | \phi_i \rangle , | \tilde{\phi}_{i,r} \rangle = \frac{\tilde{\Pi}_r | \phi_i \rangle}{\sqrt{\langle \phi_i | \tilde{\Pi}_r | \phi_i \rangle}} \right\}_{i,r}$$
(E6)

which can be thought as the ensemble obtained from the optimal ensemble $\{p_i, |\phi_i\rangle\}$ by measuring in the basis $\{\tilde{\Pi}_r\}_r$. It can be easily seen that

1. Eq.(E4) together with $\sum_i p_i |\phi_i\rangle\langle\phi_i| = \rho$ imply

$$\sum_{i,r} \tilde{p}_{i,r} |\tilde{\phi}_{i,r}\rangle \langle \tilde{\phi}_{i,r}| = \rho . \tag{E7}$$

2. The average variance for the ensemble $\{\tilde{p}_{i,r}, |\tilde{\phi}_{i,r}\rangle\}_{i,r}$ satisfies

$$\sum_{i,r} \tilde{p}_{i,r} V_H(|\tilde{\phi}_{i,r}\rangle) = \sum_i p_i \sum_r \langle \phi_i | \tilde{\Pi}_r | \phi_i \rangle V_H(|\tilde{\phi}_{i,r}\rangle)$$
 (E8)

$$\leq \sum_{i} p_{i} V_{H}(|\phi_{i}\rangle) \tag{E9}$$

$$=F_H(\rho),\tag{E10}$$

where to get the second line we have used the fact that variance is a concave function and the last line follows from Eq.(E5). But, from theorem 19 we know that this average cannot be less than QFI (This can also be seen directly from convexity of QFI). Therefore, we conclude that

$$\sum_{i,r} \tilde{p}_{i,r} V_H(|\tilde{\phi}_{i,r}\rangle) = F_H(\rho). \tag{E11}$$

- 3. Since for each subspace $\tilde{\Pi}_r$ the difference between any two energy level is an integer multiple of $2\pi/\tau$, for any i and r the period of state $\frac{\tilde{\Pi}_r|\phi_i\rangle}{\sqrt{\langle\phi_i|\tilde{\Pi}_r|\phi_i\rangle}}$ is $\tau_i=\tau/k_i$ for an integer k_i , or is zero, i.e. $\frac{\tilde{\Pi}_r|\phi_i\rangle}{\sqrt{\langle\phi_i|\tilde{\Pi}_r|\phi_i\rangle}}$ is an energy eigenstate.
- 4. Since for any time $t < \tau$, $e^{-iHt}\rho e^{iHt} \neq \rho$, we conclude that for any $t < \tau$ there should be at least one pure state $|\tilde{\phi}_{i,r}\rangle$ such that $|\langle \tilde{\phi}_{i,r}|e^{-iHt}|\tilde{\phi}_{i,r}\rangle| \neq 1$. Equivalently, this means that the greatest common divisors of integers $k_i = \tau/\tau_i$ is one.

Next, using the results of Sec. A on pure state transformations, we can prove the following lemma.

Lemma 25. For a given Hamiltonian H and a finite set of pure states $S = \{|\psi_i\rangle\}_i$, suppose all states are periodic with a period which is an integer fraction of τ , i.e. $|\langle \psi_i|e^{-iH\tau}|\psi_i\rangle|=1$. Furthermore, suppose for any $t<\tau$ there is, at least, one state $|\psi_i\rangle$ such that $|\langle \psi_i|e^{-iHt}|\psi_i\rangle|<1$. For any state $|\psi_i\rangle$ let $r_i>0$ be an arbitrary positive number. Then,

1. In the limit m goes to infinity, the energy distribution for state

$$|\Psi(m)\rangle = \bigotimes_{i \in S} |\psi_i\rangle^{\otimes \lceil r_i m \rceil} .$$
 (E12)

converges, in total variation distance, to a translated Poisson distribution with variance $m \sum_i r_i V_H(\psi_i)$.

2. Let $R > \sum_i r_i V_H(\psi_i)/V_{c-bit} = \sum_i r_i V_H(\psi_i) \tau^2/\pi^2$, where $V_{c-bit} = (\pi/\tau)^2$ is the energy variance of a system with state $|\Phi_{c-bit}\rangle$ and Hamiltonian $\pi\sigma_z/\tau$. Then, for any m there exists a TI operation which converts $\lceil mR \rceil$ copies of system with state $|\Phi_{c-bit}\rangle$ and Hamiltonian $H_{c-bit} = \pi\sigma_z/\tau$ to state $|\Psi(m)\rangle$ with an error ϵ which vanishes in the limit m goes to infinity, i.e.

$$R > \sum_{i} r_{i} V_{H}(\psi_{i}) / V_{c-bit} \rightarrow \Phi_{c-bit}^{\otimes \lceil Rm \rceil} \xrightarrow{TI} \stackrel{\epsilon}{\Longrightarrow} |\Psi(m)\rangle = \bigotimes_{i \in S} |\psi_{i}\rangle^{\otimes \lceil r_{i}m \rceil} \quad as \ m \to \infty, \epsilon \to 0 \ . \tag{E13}$$

Proof. First, we temporarily assume constants r_i are all rational numbers and show that the result follows from theorem 15: Since the set of states S has finite elements, there exists a finite integer M, such that Mr_i is an integer for all i. Now consider state

$$|\Psi(M)\rangle = \bigotimes_{i \in \mathcal{S}} |\psi_i\rangle^{\otimes (Mr_i)}.$$
 (E14)

This state has variance

$$M\sum_{i} r_i V_H(\psi_i) . {(E15)}$$

Furthermore, because for all states ψ_i , $|\langle \psi_i | e^{-iH\tau} | \psi_i \rangle| = 1$ and for any $t < \tau$ there is, at least, one state $|\psi_i\rangle$ such that $|\langle \psi_i | e^{-iHt} | \psi_i \rangle| < 1$, we conclude that the period of state $|\tilde{\Psi}(M)\rangle$ is τ , i.e.

$$\tau = \inf_{t} \{t > 0 | \left| \langle \Psi(M) | [\underbrace{e^{-iHt} \otimes \dots \otimes e^{-iHt}}_{i \text{ times}}] | \Psi(M) \rangle \right| = 1 \}.$$
 (E16)

Therefore, we can apply the results of section A, and in particular, theorem 15 which implies the transformation in Eq.(E13) is possible with TI operations, and vanishing error.

This proves the result for the special case where $\{r_i\}_i$ are rational numbers. To prove the result for general positive numbers r_i , which can include irrational numbers, for each r_i we choose a rational number $\tilde{r}_i \geq r_i$. Then, we can apply the above argument to show that the transformation is possible with any rate $R > \sum_i \tilde{r}_i V_H(\psi_i)$. But, we can choose rational number \tilde{r}_i arbitrary close to r_i . We conclude that for any $R > \sum_i \tilde{r}_i V_H(\psi_i)$ the transformation in Eq.(E13) can be implemented by a TI operation with an error which vanishes in the limit m goes to infinity. This completes the proof.

1. Proof of theorem 22

Next, we use corollary 24 and lemma 25 to prove theorem 22. The argument is similar to the argument of [106] and [6] on entanglement and coherence cost (w.r.t. Incoherent Operations).

Let $\rho = \sum_{i \in \mathcal{S}} p_i |\psi_i\rangle \langle \psi_i|$ be the optimal decomposition in corollary 24. We call \mathcal{S} the set of *alphabets*. From the proof of corollary 24 and theorem 19 it can be easily seen that this set has finite elements (assuming the Hilbert space is finite dimensional). According to corollary 24 the ensemble $\{p_i, |\psi_i\rangle\}$ can be chosen such that

1.
$$F_H(\rho) = \sum_i p_i F_H(\psi_i) = 4 \times \sum_i p_i V_H(\psi_i)$$
.

2. For all i, $|\langle \psi_i | e^{-iH\tau} | \psi_i \rangle| = 1$, and for any $t < \tau$, there is at least one state $|\psi_i\rangle$ such that $|\langle \psi_i | e^{-iHt} | \psi_i \rangle| < 1$.

Consider m copies of ρ , i.e. state

$$\rho^{\otimes m} = \sum_{\mathbf{i}} p_{\mathbf{i}} |\psi_{\mathbf{i}}\rangle \langle \psi_{\mathbf{i}}| , \qquad (E17)$$

where $\mathbf{i} = i_1 \cdots i_m$, $p_{\mathbf{i}} = p_{i_1} \cdots p_{i_m}$ and $|\psi_{\mathbf{i}}\rangle = |\psi_{i_1}\rangle \otimes \cdots \otimes |\psi_{i_m}\rangle$.

Next, we define the set of *typical* strings: For any alphabet $l \in \mathcal{S}$, let $n_l(\mathbf{i})$ be the number of occurrence of alphabet l in the string $\mathbf{i} = i_1 \cdots i_m$. Then, for any $\delta > 0$, we define the set of δ -typical strings as

$$\mathcal{T}_{\delta} = \{ \mathbf{i} = i_1 \cdots i_m | \forall l \in \mathcal{S} : \left| \frac{n_l(\mathbf{i})}{m} - p_l \right| \le \delta \}.$$
 (E18)

In other words, \mathcal{T}_{δ} is the set of all strings for which the relative frequency of any alphabet l is between $p_l - \delta$ and $p_l + \delta$. Now consider the decomposition of state $\rho^{\otimes m}$ as

$$\rho^{\otimes m} = \sum_{\mathbf{i}} p_{\mathbf{i}} |\psi_{\mathbf{i}}\rangle \langle \psi_{\mathbf{i}}| = \sum_{\mathbf{i} \in \mathcal{T}_{\delta}} p_{\mathbf{i}} |\psi_{\mathbf{i}}\rangle \langle \psi_{\mathbf{i}}| + \sum_{\mathbf{j} \notin \mathcal{T}_{\delta}} p_{\mathbf{i}} |\psi_{\mathbf{i}}\rangle \langle \psi_{\mathbf{i}}|.$$
 (E19)

Suppose for any string i in the typical set \mathcal{T}_{δ} we prepare state $|\psi_i\rangle$ with probability p_i . Furthermore, with probability

$$p_{\text{err}} = \sum_{\mathbf{i} \notin \mathcal{T}_s} p_{\mathbf{i}} = 1 - \sum_{\mathbf{i} \in \mathcal{T}_s} p_{\mathbf{i}} , \qquad (E20)$$

we prepare a fixed time-invariant state σ_{inv} , e.g. the totally mixed state. The resulting state is

$$\tilde{\rho}_m = \sum_{\mathbf{i} \in \mathcal{T}_{\delta}} p_{\mathbf{i}} |\psi_{\mathbf{i}}\rangle \langle \psi_{\mathbf{i}}| + p_{\text{err}} \sigma_{\text{inv}}. \tag{E21}$$

Then, using the standard typicality arguments we can show that for any fixed $\delta > 0$, in the limit m goes to infinity, almost all strings, except a vanishing fraction of them, are in the typical set \mathcal{T}_{δ} , and therefore p_e goes to zero. It follows that in the limit m goes to infinity the trace distance between this state and the desired state $\rho^{\otimes m}$ vanishes, i.e.

$$\|\rho^{\otimes m} - \tilde{\rho}_m\|_1 \to 0. \tag{E22}$$

Next, we argue that for any typical string \mathbf{i} in the typical set \mathcal{T}_{δ} state $|\psi_{\mathbf{i}}\rangle$ can be prepared using TI operations by consuming approximately $\lceil Rm \rceil$ copies of state $\Phi_{\text{c-bit}}$ for $R > \sum_i p_i V_H(\psi_i)/V_{c-bit}$. First, note that, up to a permutation, state $|\psi_{\mathbf{i}}\rangle = |\psi_{i_1}\rangle \otimes \cdots \otimes |\psi_{i_m}\rangle$ can be written as

$$|\Psi_{\mathbf{i}}\rangle = \bigotimes_{i \in \mathcal{S}} |\psi_i\rangle^{\otimes n_l(\mathbf{i})} ,$$
 (E23)

where, the typicality of string i implies

$$n_l(\mathbf{i}) \le m(p_l + \delta) \ . \tag{E24}$$

Then, it follows from lemma 25 that for any

$$R > \sum_{l} (p_l + \delta) V_H(\psi_l) / V_{c-bit} , \qquad (E25)$$

and any typical string i, there exists a TI operation which implements the transformation

$$\Phi_{\text{c-bit}}^{\otimes \lceil Rm \rceil} \xrightarrow{TI} \stackrel{\epsilon}{\approx} |\Psi_{\mathbf{i}}\rangle \quad \text{as } m \to \infty, \epsilon \to 0 \ . \tag{E26}$$

We conclude that for any $\delta>0$ and any $R>\sum_l(p_l+\delta)V_H(\psi_l)/V_{c-bit}$ there is a TI operation such that

$$\Phi_{\text{c-bit}}^{\otimes \lceil Rm \rceil} \xrightarrow{TI} \stackrel{\epsilon}{\approx} \tilde{\rho}_m = \sum_{\mathbf{i}} p_{\mathbf{i}} |\psi_{\mathbf{i}}\rangle \langle \psi_{\mathbf{i}}| + p_{\text{err}} \sigma_{\text{inv}} \text{ as } m \to \infty, \epsilon \to 0 .$$
 (E27)

Combining this with Eq.(E22), which implies as $m \to \infty$ the trace distance $\|\tilde{\rho}_m - \rho^{\otimes m}\|_1 \to 0$, and using the fact that δ can be chosen arbitrarily small, we find that for any $R > \sum_l p_l V_H(\psi_l)/V_{c-bit}$

$$\Phi_{\text{c-bit}}^{\otimes \lceil Rm \rceil} \xrightarrow{TI} \stackrel{\epsilon}{\approx} \rho^{\otimes m} \text{ as } m \to \infty, \epsilon \to 0 ,$$
 (E28)

which completes the proof of $C_c^{\rm TI}(\rho) \leq F_H(\rho)/F_{\text{c-bit}}$. The proof of the second part, i.e. $C_c^{\rm TI}(\rho) \geq F_H(\rho)/F_{\text{c-bit}}$ follows from

theorem 20 in Sec.(D) which implies Quantum Fisher Information cannot increase in the iid regime.

Appendix F: Purity of Coherence

In this paper we introduce a new measure of asymmetry, which we call it Purity of Coherence. The purity of coherence, with respect to the eigenbasis of an observable H, is defined by

$$P_H(\rho) \equiv \text{Tr}(H\rho^2 H \rho^{-1}) - \text{Tr}(\rho H^2) , \qquad (F1)$$

$$= f_H(\rho) - f_H(\mathcal{D}(\rho)) \tag{F2}$$

if $\sup(H\rho H)\subseteq \sup(\rho)$, and $P_H(\rho)=\infty$ otherwise. Here, $f_H(\rho)\equiv \operatorname{Tr}(H\rho^2H\rho^{-1})$, and

$$\mathcal{D}(\rho) = \sum_{n} P_n \rho P_n = \lim_{T \to \infty} \frac{1}{T} \int_0^T dt \ e^{-iHt} \rho e^{iHt} \ , \tag{F3}$$

is the map that dephases ρ in the eigen-basis of $H=\sum_n E_n P_n$ (also known as the resource-destroying map [107]).

Using the spectral decomposition of state ρ , as $\rho=\sum_j p_j|\psi_j\rangle\langle\psi_j|$, we can rewrite the formula for the purity of coherence as

$$P_H(\rho) = \sum_{j,k} \frac{p_k^2 - p_j^2}{p_j} |\langle \psi_k | H | \psi_j \rangle|^2.$$
 (F4)

1. Properties of purity of coherence

The important properties of purity of coherence, such as monotonicity under TI operations and convexity, follow from the properties of the function

$$\overline{Q}_2(\rho \| \sigma) \equiv \text{Tr}(\rho^2 \sigma^{-1}),\tag{F5}$$

if $\sup(\rho) \subseteq \sup(\sigma)$, and $\overline{Q}_2(\rho \| \sigma) = \infty$ otherwise (As we will discuss later, the logarithm of this function is the (Petz) relative Renyi entropy for $\alpha = 2$).

In particular, this function satisfies the following properties (See [78] for further discussions and proofs of these properties):

- Unitary invariance: It is invariant under any unitary transformation U, i.e. $\overline{Q}_2(U\rho U^\dagger \| U\sigma U^\dagger) = \overline{Q}_2(\rho \| \sigma)$.
- Joint convexity: For any $0 \le p \le 1$:

$$p \, \overline{Q}_2(\rho_1 \| \sigma_1) + (1 - p) \overline{Q}_2(\rho_2 \| \sigma_2) \ge \overline{Q}_2 \left([p\rho_1 + (1 - p)\rho_2] \middle\| [p\sigma_1 + (1 - p)\sigma_2] \right)$$
 (F6)

ullet Information-processing inequality: For any Completely Positive Trace-Preserving map \mathcal{E} ,

$$\overline{Q}_2(\mathcal{E}(\rho)||\mathcal{E}(\sigma)) \le \overline{Q}_2(\rho||\sigma) . \tag{F7}$$

This follows from the Steinspring dilation theorem, together with unitary invariance and joint convexity.

Using these properties, it can be easily seen that for any time t, the function $B(\rho) \equiv \overline{Q}_2(\rho \| e^{-itH} \rho e^{itH})$ is monotone under any TI operation \mathcal{E}_{TI} :

$$B(\rho) = \overline{Q}_2(\rho \| e^{-itH} \rho e^{itH}) \ge \overline{Q}_2(\mathcal{E}_{\Pi}(\rho) \| \mathcal{E}_{\Pi}(e^{-itH} \rho e^{itH}))$$
 (F8)

$$= \overline{Q}_2(\mathcal{E}_{\text{TI}}(\rho) \| e^{-itH} \mathcal{E}_{\text{TI}}(\rho) e^{itH})$$
 (F9)

$$= B(\mathcal{E}_{TI}(\rho)) , \qquad (F10)$$

where the inequality follows from the information processing inequality for \overline{Q}_2 , and the second line follows from the fact that \mathcal{E}_{TI} is a TI operation.

The connection between this function and the purity of coherence follows from the fact that for small Δt ,

$$\overline{Q}_{2}(\rho \| e^{-i\Delta t H} \rho e^{i\Delta t H}) = \text{Tr}(\rho^{2} (e^{-i\Delta t H} \rho e^{i\Delta t H})^{-1})$$

$$= \text{Tr}(\rho^{2} e^{-i\Delta t H} \rho^{-1} e^{i\Delta t H})$$

$$= \left[1 + \Delta t^{2} \text{Tr}(\rho^{2} H \rho^{-1} H) - \Delta t^{2} \text{Tr}(\rho^{2} H^{2} \rho^{-1})/2 - \Delta t^{2} \text{Tr}(\rho^{2} \rho^{-1} H^{2})/2 + \mathcal{O}(\Delta t^{4}) \right]$$

$$= 1 + \Delta t^{2} P_{H}(\rho) + \mathcal{O}(\Delta t^{4}) .$$
(F13)

In other words, $P_H(\rho)$ is 2 time the second derivative of function $\overline{Q}_2(\rho \| e^{-itH} \rho e^{itH})$ with respect to the parameter t, at t=0. Then, it follows from the joint convexity of \overline{Q}_2 in Eq.(F6), and information processing inequality in Eq.(F7), that function P_H is

• Convex: For any $0 \le p \le 1$, and any pair of states ρ_1, ρ_2 :

$$pP_H(\rho_1) + (1-p)P_H(\rho_2) \ge P_H([p\rho_1 + (1-p)\rho_2])$$
 (F14)

• Monotone: For any TI operation \mathcal{E}_{TI}

$$P_H(\mathcal{E}_{TI}(\rho)) \le P_H(\rho).$$
 (F15)

Furthermore, it turns out that the purity of coherence has the following useful properties:

- Additive: For a composite non-interacting system with the total Hamiltonian $H_{\text{tot}} = H_1 \otimes I_2 + I_1 \otimes H_2$, the purity of coherence is additive for uncorrelated states, i.e. $F_{H_{\text{tot}}}(\rho_1 \otimes \rho_2) = P_{H_1}(\rho_1) + P_{H_2}(\rho_2)$.
- Faithful: It is non-negative, and is zero if, and only if, state is incoherent, i.e. diagonal in the energy eigenbasis. To see this consider the spectral decomposition of state ρ as $\rho = \sum_j p_j |\psi_j\rangle \langle \psi_j|$. Then, we obtain the formula

$$P_{H}(\rho) = \sum_{j,k} \frac{p_k^2 - p_j^2}{p_j} |\langle \psi_k | H | \psi_j \rangle|^2 ,$$
 (F16)

$$= \sum_{j,k} \frac{(p_j - p_k)^2}{2p_j p_k} (p_j + p_k) |\langle \psi_k | H | \psi_j \rangle|^2 ,$$
 (F17)

where to get the second line we have used the fact that $|\langle \psi_k | H | \psi_j \rangle|^2$ is symmetric with respect to k and j. The last line immediately implies that $P_H(\rho) \geq 0$. Furthermore, since all the terms in the summation are non-negative, the sum will be zero iff all the individual terms are zero. That is $(p_j - p_k)^2 |\langle \psi_k | H | \psi_j \rangle|^2 = 0$ for all j, k, or equivalently $(p_j - p_k)\langle \psi_k | H | \psi_j \rangle = 0$. Multiplying both sides in $|\psi_k\rangle\langle\psi_j|$, and summing over j and k this implies $[\rho, H] = 0$ and completes the proof.

2. States with infinite purity of coherence

The definition of the purity of coherence immediately implies that for any pure state which is not an eigenstate of Hamiltonian the purity of coherence is infinite. This unboundedness, reflects the fact that given any finite copies of a generic mixed state (with full-rank density operator) it is impossible to create a single copy such pure states using TI operations. More generally,

Lemma 26. For a bounded Hamiltonian H (i.e. $||H||_{\infty} < \infty$) the purity of coherence $P_H(\rho) < \infty$ if and only if $[\Pi_{\rho}, H] = 0$, where Π_{ρ} is the projector to the support of ρ . In particular, the purity of coherence is bounded for states with full rank.

Proof. First, note that if $\sup(H\rho H)$ is not contained in the $\sup(\rho)$ then $P_H(\rho)=\infty$. But, the support of the positive operator $H\rho H$ is equal to the support of $H\Pi_\rho H$. Let $Q_\rho=I-\Pi_\rho$ be the projector operator to the kernel of ρ . Then, the support of $H\Pi_\rho H$ is contained in the support of ρ if and only if $Q_\rho H\Pi_\rho HQ_\rho=0$, or equivalently, $Q_\rho H\Pi_\rho=0$ which means $[H,\Pi_\rho]=0$. On the other hand, if $[H,\Pi_\rho]=0$ then it can be easily seen that $\mathrm{Tr}(H\rho^2 H\rho^{-1})=\mathrm{Tr}(H\rho^2 H\Pi_\rho\rho^{-1}\Pi_\rho)<\infty$ and therefore $P_H(\rho)$ is finite, provided that H is bounded.

The following proposition follows immediately from this lemma together with the monotonicity of the purity of coherence under TI operations.

Remark 27. Suppose under a TI operation the input state ρ is transformed to the output state σ . Let Π_{ρ} and Π_{σ} be the projectors to the supports of ρ and σ , respectively. If $[\Pi_{\rho}, H_{in}] = 0$ then $[\Pi_{\sigma}, H_{in}] = 0$, where $H_{in/out}$ are the input/output (bounded) Hamiltonians.

In the following, we present an interpretation and a different proof of this result in terms of the notion of unambiguous state discrimination [108], which clarifies the physical relevance of condition $[H, \Pi_{\rho}] = 0$.

Recall that two density operators can be unambiguously discriminated with a non-zero probability iff their supports are not identical [108]. The support of state $e^{-iHt}\rho e^{iHt}$ is $e^{-iHt}\Pi_{\rho}e^{iHt}$, which is equal to Π_{ρ} for all $t\in\mathbb{R}$, if and only if $[H,\Pi_{\rho}]=0$. Therefore, we conclude that for some $t\in\mathbb{R}$ two states $e^{-iHt}\rho e^{iHt}$ and ρ can be unambiguously discriminated, with a non-zero probability, iff $[H,\Pi_{\rho}]\neq 0$.

Next, we note that if the probability of unambiguous discrimination of two states ρ_1 and ρ_2 is zero, then this probability remains zero under any Completely Positive Trace Preserving map \mathcal{E} , i.e. two states $\sigma_1 = \mathcal{E}(\rho_1)$ and $\sigma_2 = \mathcal{E}(\rho_2)$ will also have the same support. This immediately implies that if the probability of unambiguous discrimination of ρ and $e^{-iHt}\rho e^{iHt}$ is zero, then for any TI operation \mathcal{E}_{TI} , the probability of unambiguous discrimination of the two states $\sigma = \mathcal{E}_{\text{TI}}(\rho)$ and

$$\mathcal{E}_{\text{TI}}(e^{-iHt}\rho e^{iHt}) = e^{-iHt}\mathcal{E}_{\text{TI}}(\rho)e^{iHt} = e^{-iHt}\sigma e^{iHt}$$
(F18)

should also be zero. We conclude that if $[\Pi_{\rho}, H] = 0$, and $\sigma = \mathcal{E}_{TI}(\rho)$ for a TI operation \mathcal{E}_{TI} , then $[\Pi_{\sigma}, H] = 0$.

3. Connection with (Petz) relative Renyi entropy

Similar to QFI, function $P_H(\rho)$ also determines how fast state ρ becomes distinguishable from its time evolved version $e^{-iHt}\rho e^{iHt}$ and is closely related to the (Petz) relative Renyi entropies. For $\alpha \in (0,1) \cup (1,2]$ The (Petz) relative Renyi entropy for is defined as

$$D_{\alpha}(\rho \| \sigma) = \frac{1}{\alpha - 1} \log \operatorname{Tr}(\rho^{\alpha} \sigma^{1 - \alpha}), \qquad \operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma), \text{ or } \alpha \in (0, 1)$$
 (F19a)

$$D_{\alpha}(\rho \| \sigma) = \infty,$$
 otherwise (F19b)

Note that the in the special case of $\alpha = 2$, we have $D_2(\rho \| \sigma) = \log \overline{Q}_2(\rho \| \sigma)$.

The relative Renyi entropy can be interpreted as a measure of distinguishability of states. In particular, it is non-negative and $D_{\alpha}(\rho \| \sigma)$ is zero if and only if $\rho = \sigma$. Furthermore, it satisfies information processing inequality for $\alpha \in [0, 2]/\{1\}$ [77, 78], that is for any Completely Positive Trace-Preserving (CPTP) map \mathcal{E} , it holds that $D_{\alpha}(\rho \| \sigma) \geq D_{\alpha}(\mathcal{E}(\rho) \| \mathcal{E}(\sigma))$.

It can be easily seen that for small time Δt ,

$$D_{\alpha}(\rho \| e^{-i\Delta t H} \rho e^{i\Delta t H}) = \frac{1}{\alpha - 1} \Delta t^2 [\text{Tr}(\rho^{\alpha} H \rho^{1 - \alpha} H) - \text{Tr}(\rho H^2)] + \mathcal{O}(\Delta t^4) . \tag{F20}$$

Then, using the arguments we used in the case of the purity of coherence, we can see that all functions in the family

$$\operatorname{Tr}(\rho^{\alpha}H\rho^{1-\alpha}H) - \operatorname{Tr}(\rho H^2), \qquad 1 < \alpha \le 2$$
 (F21)

satisfies all the essential properties of the purity of coherence, such as monotonicity under TI operations, additivity, and convexity.

However, the reason that in this paper we focus on the case of $\alpha=2$, and call it the *purity of coherence*, is that as a mixed state ρ converges to a pure state, higher value of α implies faster divergence of the function. Therefore, to capture the unreachability of pure coherent states from mixed states, we focus on the case of $\alpha=2$, which yields the fastest divergence for pure states, and hence the strongest bound (It turns out that for $\alpha>2$, the (Petz) relative Renyi entropy does not satisfy information processing inequality).

4. Purity of coherence is lower-bounded by Quantum Fisher Information

In this section we show that the purity of coherence is lower-bounded by the Quantum Fisher Information (QFI). That is for any state ρ , and Hamiltonian H, $P_H(\rho) \ge F_H(\rho)$.

For state ρ with the spectral decomposition $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$, the QFI is given by

$$F_H(\rho) = 2\sum_{k,l} \frac{(p_k - p_l)^2}{p_k + p_l} |\langle \psi_k | H | \psi_l \rangle|^2 ,$$
 (F22)

Using the definition $P_H(\rho) = \text{Tr}(H\rho^2H\rho^{-1}) - \text{Tr}(\rho H^2)$ and the spectral decomposition $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$, we can find a similar formula for the purity of coherence,

$$P_H(\rho) = \sum_{k,l} \frac{p_k^2 - p_l^2}{p_l} |\langle \psi_k | H | \psi_l \rangle|^2 , \qquad (F23)$$

which can be rewritten as

$$P_H(\rho) = \sum_{k,l} \frac{p_k^2 - p_l^2}{p_l} |\langle \psi_k | H | \psi_l \rangle|^2$$
 (F24)

$$= \frac{1}{2} \sum_{k,l} \left(\frac{p_k^2 - p_l^2}{p_l} + \frac{p_l^2 - p_k^2}{p_k} \right) |\langle \psi_k | H | \psi_l \rangle|^2$$
 (F25)

$$= \sum_{k,l} \frac{p_k(p_k^2 - p_l^2) + p_l(p_l^2 - p_k^2)}{2p_l p_k} |\langle \psi_k | H | \psi_l \rangle|^2$$
 (F26)

$$= \sum_{k,l} \frac{(p_k - p_l)^2 (p_k + p_l)}{2p_l p_k} |\langle \psi_k | H | \psi_l \rangle|^2,$$
 (F27)

where we have used the fact that $|\langle \psi_k | H | \psi_l \rangle|^2$ is symmetric with respect to k and l. Therefore,

$$P_{H}(\rho) = \sum_{k,l} \frac{(p_k^2 - p_l^2)^2}{2p_l p_k (p_l + p_k)} |\langle \psi_k | H | \psi_l \rangle|^2 .$$
 (F28)

Comparing this with Eq.(F22) for QFI we find

$$P_H(\rho) - F_H(\rho) = \sum_{k,l} \frac{(p_k^2 - p_l^2)^2 - 4p_k p_l (p_k - p_l)^2}{2p_l p_k (p_l + p_k)} |\langle \psi_k | H | \psi_l \rangle|^2$$
 (F29)

$$= \sum_{k,l} \frac{p_k^4 - 2p_k^2 p_l^2 + p_l^4 - 4p_k p_l (p_k^2 + p_l^2 - 2p_k p_l)}{2p_l p_k (p_l + p_k)} |\langle \psi_k | H | \psi_l \rangle|^2$$
 (F30)

$$= \sum_{k,l} \frac{p_k^4 + 6p_k^2 p_l^2 + p_l^4 - 4p_k^3 p_l - 4p_l^3 p_k}{2p_l p_k (p_l + p_k)} |\langle \psi_k | H | \psi_l \rangle|^2$$
 (F31)

$$= \sum_{k,l} \frac{(p_k - p_l)^4}{2p_l p_k (p_l + p_k)} |\langle \psi_k | H | \psi_l \rangle|^2$$
 (F32)

$$\geq 0$$
. (F33)

Therefore, $P_H(\rho) \geq F_H(\rho)$.

5. Purity of coherence for Qubits

Consider a general qubit state ρ with the spectral decomposition

$$\rho = p|\psi\rangle\langle\psi| + (1-p)|\psi^{\perp}\rangle\langle\psi^{\perp}| \tag{F34}$$

where $\langle \psi^{\perp} | \psi \rangle = 0$. Then,

$$\rho^{-1} = \frac{1}{p} |\psi\rangle\langle\psi| + \frac{1}{1-p} |\psi^{\perp}\rangle\langle\psi^{\perp}|, \qquad (F35)$$

and

$$\rho^2 = p^2 |\psi\rangle\langle\psi| + (1-p)^2 |\psi^{\perp}\rangle\langle\psi^{\perp}|. \tag{F36}$$

This implies

$$\operatorname{Tr}(\rho^{2}H\rho^{-1}H) = p\operatorname{Tr}(H|\psi\rangle\langle\psi|H|\psi\rangle\langle\psi|) + \frac{p^{2}}{1-p}\operatorname{Tr}(H|\psi\rangle\langle\psi|H|\psi^{\perp}\rangle\langle\psi^{\perp}|)$$
 (F37)

$$+\frac{(1-p)^2}{p}\operatorname{Tr}(H|\psi^{\perp}\rangle\langle\psi^{\perp}|H|\psi\rangle\langle\psi|) + (1-p)\operatorname{Tr}(H|\psi^{\perp}\rangle\langle\psi^{\perp}|H|\psi^{\perp}\rangle\langle\psi^{\perp}|). \tag{F38}$$

Using the fact that

$$Tr(H|\psi\rangle\langle\psi|H|\psi^{\perp}\rangle\langle\psi^{\perp}|) = V_H(\psi) = V_H(\psi^{\perp}), \qquad (F39)$$

we find

$$Tr(\rho^{2}H\rho^{-1}H) = \left(\frac{p^{2}}{1-p} + \frac{(1-p)^{2}}{p}\right) \times V(\psi) + p|\langle\psi|H|\psi\rangle|^{2} + (1-p)|\langle\psi^{\perp}|H|\psi^{\perp}\rangle|^{2}.$$
 (F40)

Then, we find

$$P_H(\rho) = \text{Tr}(\rho^2 H \rho^{-1} H) - \text{Tr}(\rho H^2) = \left(\frac{p^2}{1-p} + \frac{(1-p)^2}{p} - 1\right) \times V(\psi)$$
 (F41)

$$= \frac{(1-2p)^2}{p(1-p)} \times V(\psi) . \tag{F42}$$

Next, using the formula for Quantum Fisher information for the family of states $e^{-iHt}\rho e^{iHt}$ with parameter t,

$$F_H(\rho) = 2\sum_{i,j} \frac{(p_i - p_j)^2}{p_i + p_j} |\langle \psi_i | H_S | \psi_j \rangle|^2 .$$
 (F43)

where $\sum_i p_i |\psi_i\rangle\langle\psi_i|$ is the spectral decomposition of ρ . Applying this to $\rho=p|\psi\rangle\langle\psi|+(1-p)|\psi^{\perp}\rangle\langle\psi^{\perp}|$, we find

$$F_H(\rho) = 4(1-2p)^2 \times V_H(\psi).$$
 (F44)

Therefore,

$$P_H(\rho) = \frac{F_H(\rho)}{4p(1-p)}$$
 (F45)

Finally, note that

$$1 - \text{Tr}(\rho^2) = 1 - [p^2 + (1 - p)^2] = 2p - 2p^2 = 2p(1 - p).$$
 (F46)

Therefore,

$$P_H(\rho) = \frac{F_H(\rho)}{2[1 - \text{Tr}(\rho^2)]}$$
 (F47)

6. Purity of coherence for states close to the totally mixed state

Any general state ρ whose trace distance from the totally mixed state is $\|\rho - I/d\|_1 = \epsilon \ge 0$ can be written as

$$\rho = \frac{I}{d} + \epsilon A \,, \tag{F48}$$

where A is a Hermitian operator with Tr(A) = 0 and $||A||_1 = 1$.

In the following we calculate $P_H(\rho)$ and $F_H(\rho)$ in the limit of $\epsilon \ll 1$.

First, recall that

$$P_H(\rho) = Tr(H\rho^2 H \rho^{-1}) - Tr(\rho H^2). \tag{F49}$$

For $\rho = \frac{I}{d} + \epsilon A$ we find

$$Tr(\rho H^2) = \frac{1}{d}Tr(H^2) + \epsilon Tr(H^2 A)$$
 (F50)

Then.

$$\operatorname{Tr}(H\rho^2 H \rho^{-1}) = \operatorname{Tr}(H\left[\frac{I}{d^2} + \frac{2}{d}\epsilon A + \epsilon^2 A^2\right] H\left[\frac{I}{d} + \epsilon A\right]^{-1}) \tag{F51}$$

$$= d\text{Tr}(H\left[\frac{I}{d^2} + \frac{2}{d}\epsilon A + \epsilon^2 A^2\right]H\frac{1}{I + \epsilon dA})$$
 (F52)

$$= d\text{Tr}(H\left[\frac{I}{d^2} + \frac{2}{d}\epsilon A + \epsilon^2 A^2\right]H\left[I - \epsilon dA + (\epsilon dA)^2 + \mathcal{O}(\epsilon^3)\right])$$
 (F53)

$$= \frac{1}{d}\operatorname{Tr}(H^2) + d\operatorname{Tr}(H[\frac{2}{d}\epsilon A]HI) + d\operatorname{Tr}(H[\frac{I}{d^2}]H[-\epsilon dA])$$
 (F54)

$$+ d\operatorname{Tr}(H[\frac{2}{d}\epsilon A]H[-\epsilon dA]) + d\operatorname{Tr}(H[\epsilon^2 A^2]HI) + d\operatorname{Tr}(H[\frac{I}{d^2}]H[(\epsilon dA)^2]) + \mathcal{O}(\epsilon^3)$$
 (F55)

$$= \frac{1}{d} \operatorname{Tr}(H^2) + \epsilon \operatorname{Tr}(H^2 A) + d2\epsilon^2 [\operatorname{Tr}(H^2 A^2) - \operatorname{Tr}(H A H A)] + \mathcal{O}(\epsilon^3) . \tag{F56}$$

Therefore

$$P_H(\rho) = Tr(H\rho^2 H \rho^{-1}) - Tr(\rho H^2) = \epsilon^2 2d[Tr(H^2 A^2) - 2Tr(HAHA)] + \mathcal{O}(\epsilon^3). \tag{F57}$$

Next, we calculate Quantum Fisher Information for this state. Recall the formula

$$F_H(\rho) = 2\sum_{i,j} \frac{(p_i - p_j)^2}{p_i + p_j} |\langle \psi_i | H_S | \psi_j \rangle|^2 .$$
 (F58)

Let $A = \sum_i a_i |i\rangle\langle i|$ be the spectral decomposition of A. Then,

$$\rho = \sum_{i} (\epsilon a_i + \frac{1}{d}) |i\rangle\langle i| . \tag{F59}$$

This implies

$$F_H(\rho) = 2\sum_{i,j} \frac{(p_i - p_j)^2}{p_i + p_j} |\langle \psi_i | H_S | \psi_j \rangle|^2 .$$
 (F60)

For $\rho = \frac{I}{d} + \epsilon A$ we find

$$F_H(\rho) = 2\sum_{i,j} \frac{\epsilon^2 (a_i - a_j)^2}{2/d + \epsilon (a_i + a_j)} |\langle i|H|j\rangle|^2.$$
 (F61)

Expanding this we find

$$F_H(\rho) = d\epsilon^2 \sum_{i,j} \frac{(a_i - a_j)^2}{1 + \epsilon d(a_i + a_j)/2} |\langle i|H|j\rangle|^2$$
(F62)

$$= d\epsilon^2 \sum_{i,j} (a_i - a_j)^2 |\langle i|H|j\rangle|^2 + \mathcal{O}(\epsilon^3)$$
 (F63)

$$= d\epsilon^2 \sum_{i,j} (a_i^2 + a_j^2 - 2a_i a_j) |\langle i|H|j\rangle|^2 + \mathcal{O}(\epsilon^3)$$
 (F64)

$$=2d\epsilon^{2}\operatorname{Tr}(A^{2}H^{2})-2d\epsilon^{2}\operatorname{Tr}(HAHA)+\mathcal{O}(\epsilon^{3})$$
(F65)

$$=2d\epsilon^{2}[\operatorname{Tr}(A^{2}H^{2})-\operatorname{Tr}(HAHA)]+\mathcal{O}(\epsilon^{3})\;. \tag{F66}$$

Comparing this with $P_H(\rho)=\epsilon^2 2d[{\rm Tr}(H^2A^2)-2{\rm Tr}(HAHA)]+\mathcal{O}(\epsilon^3)$ we find

$$\frac{P_H(\rho)}{F_H(\rho)} = 1 + \mathcal{O}(\epsilon) . \tag{F67}$$

7. Purity of coherence for a mixed state close to a pure state

In this section we find a useful bound on the purity of coherence for mixed states which are close to a pure state. This bound will be used later to study coherence distillation.

Lemma 28. Let $p_{max} = \|\sigma\|_{\infty}$ be the largest eigenvalue of σ and $|\Phi\rangle$ be the corresponding eigenvector. Then,

$$P_H(\sigma) \ge V_H(\Phi) \times \left(\frac{p_{max}^2}{1 - p_{max}} - 1\right). \tag{F68}$$

Proof. Recall that for any state ρ with the spectral decomposition $\rho = \sum_j q_j |\phi_j\rangle\langle\phi_j|$, purity of coherence is given by

$$P_H(\rho) = \sum_{j,k} \frac{q_k^2 - q_j^2}{q_j} |\langle \phi_k | H | \phi_j \rangle|^2.$$
 (F69)

For state σ , let $p_{\max} = \|\sigma\|_{\infty}$ be its largest eigenvalue and $|\Phi\rangle$ be the corresponding eigenvector. Also, let $\{p_j\}_j$ be the rest of the eigenvalues and $\{|\psi_j^{\perp}\rangle\}$ be the corresponding eigenvectors. Therefore,

$$\sigma = p_{\text{max}} |\Phi\rangle\langle\Phi| + \sum_{j} p_{j} |\psi_{j}^{\perp}\rangle\langle\psi_{j}^{\perp}| . \tag{F70}$$

Putting this into Eq.(F69) and using the fact that for any pair of j and k the sum of two terms

$$(\frac{p_k^2 - p_j^2}{p_j} + \frac{p_j^2 - p_k^2}{p_k})|\langle \psi_k^{\perp}|H|\psi_j^{\perp}\rangle|^2 = \frac{(p_j - p_k)^2}{2p_j p_k}(p_j + p_k)|\langle \psi_k^{\perp}|H|\psi_j^{\perp}\rangle|^2 \ge 0,$$
 (F71)

is non-negative, we find that

$$P_H(\sigma) \ge \sum_j (\frac{p_{\text{max}}^2 - p_j^2}{p_j} + \frac{p_j^2 - p_{\text{max}}^2}{p_{\text{max}}}) |\langle \Phi | H | \psi_j^{\perp} \rangle|^2,$$
 (F72)

where in the summation we have dropped all the terms which do not involve $|\Phi\rangle$. Then, we find

$$P_H(\sigma) \ge \sum_{i} \left(\frac{p_{\text{max}}^2 - p_j^2}{p_j} + \frac{p_j^2 - p_{\text{max}}^2}{p_{\text{max}}}\right) |\langle \Phi | H | \psi_j^{\perp} \rangle|^2$$
 (F73)

$$\geq \sum_{i} \left(\frac{p_{\text{max}}^2}{p_j} - [p_{\text{max}} + p_j]\right) |\langle \Phi | H | \psi_j^{\perp} \rangle|^2 \tag{F74}$$

$$\geq \sum_{j} \left(\frac{p_{\text{max}}^2}{p_j} - 1\right) \left| \langle \Phi | H | \psi_j^{\perp} \rangle \right|^2 \tag{F75}$$

$$\geq \left(\frac{p_{\text{max}}^2}{1 - p_{\text{max}}} - 1\right) \sum_{j} |\langle \Phi | H | \psi_j^{\perp} \rangle|^2 \tag{F76}$$

$$= \left(\frac{p_{\text{max}}^2}{1 - p_{\text{max}}} - 1\right) V_H(\Phi) , \qquad (F77)$$

where to get the third inequality we have used the fact that $p_{\text{max}} + p_j \le 1$, to get the fourth inequality we have used the fact that $p_j \le 1 - p_{\text{max}}$, and to get the last equality we have used

$$\sum_{j} |\langle \Phi | H | \psi_{j}^{\perp} \rangle|^{2} = \langle \Phi | H (\sum_{j} |\psi_{j}^{\perp} \rangle \langle \psi_{j}^{\perp}| + |\Phi \rangle \langle \Phi |) H |\Phi \rangle - \langle \Phi | H (|\Phi \rangle \langle \Phi |) H |\Phi \rangle \tag{F78}$$

$$= \langle \Phi | H^2 | \Phi \rangle - \langle \Phi | H | \Phi \rangle^2 = V_H(\Phi) . \tag{F79}$$

This completes the proof of lemma.

This lemma has the following corollary.

Corollary 29. Suppose $||\Psi\rangle\langle\Psi| - \sigma||_1 \le \epsilon$. Then, there exists a pure state Φ (namely the eigenvector of σ with the largest eigenvalue) whose overlap with Ψ is lower bounded by $|\langle\Psi|\Phi\rangle|^2 \ge 1 - \epsilon$, and satisfies

$$P_H(\sigma) \ge V_H(\Phi) \times (2\epsilon^{-1} - 3) . \tag{F80}$$

Proof. Let $|\Phi\rangle$ be the eingenvector of σ with the largest eigenvalue, denoted by p_{max} . Since the sum of the eigenvalues of $|\Psi\rangle\langle\Psi|-\sigma$ is zero (because it is traceless) and the sum of the absolute value of its eigenvalues are bounded by ϵ (because $||\Psi\rangle\langle\Psi|-\sigma||_1 \leq \epsilon$), we find

$$\||\Psi\rangle\langle\Psi| - \sigma\|_{\infty} = \max_{|\theta\rangle} \frac{|\langle\theta|[|\Psi\rangle\langle\Psi| - \sigma]|\theta\rangle|}{\langle\theta|\theta\rangle} \le \epsilon/2.$$
 (F81)

In particular this means that

$$\langle \Psi | [|\Psi\rangle\langle\Psi| - \sigma]|\Psi\rangle \le ||\Psi\rangle\langle\Psi| - \sigma||_{\infty} \le \epsilon/2 , \tag{F82}$$

and therefore,

$$1 - \epsilon/2 \le \langle \Psi | \sigma | \Psi \rangle. \tag{F83}$$

But, for any state $|\Psi\rangle$, the expectation value $\langle\Psi|\sigma|\Psi\rangle$ should be less than the maximum eigenvalue of σ , i.e. p_{max} . Therefore, we conclude that

$$p_{\text{max}} \ge 1 - \epsilon/2 \,. \tag{F84}$$

Combining this bound with lemma 28 we find

$$P_H(\sigma) \ge V_H(\Phi) \times \left[\frac{p_{\text{max}}^2}{1 - p_{\text{max}}} - 1\right]$$
 (F85)

$$\geq V_H(\Phi) \times (2\epsilon^{-1} - 3) . \tag{F86}$$

Finally, we note that because $\||\Psi\rangle\langle\Psi|-\sigma\|_{\infty}\leq\epsilon/2$, then

$$\langle \Phi | [\sigma - |\Psi\rangle\langle\Psi|] | \Phi \rangle = p_{\text{max}} - |\langle\Phi|\Psi\rangle|^2 \le \epsilon/2 , \qquad (F87)$$

which implies

$$p_{\text{max}} - \epsilon/2 \le |\langle \Phi | \Psi \rangle|^2 \,. \tag{F88}$$

Combining this with Eq.(F84) we find

$$1 - \epsilon < |\langle \Phi | \Psi \rangle|^2 . \tag{F89}$$

Appendix G: Distillable Coherence

Next, we use the purity of coherence to study asymptotic transformations and prove the following theorem, which implies that distillable coherence is zero for any state with bounded purity of coherence.

Theorem 30. Suppose for any integer n, there exists a TI operation \mathcal{E}_n which transforms n copies of system with state ρ and Hamiltonian H_1 to $\lceil Rn \rceil$ copies of a system with Hamiltonian H_2 and pure state ψ , i.e. $\rho^{\otimes n} \xrightarrow{TI} \stackrel{\epsilon_n}{\approx} \psi^{\otimes \lceil Rn \rceil}$, with an error ϵ_n which vanishes in the limit n goes to infinity, such that

$$\left\| \mathcal{E}_n(\rho^{\otimes n}) - \psi^{\otimes \lceil Rn \rceil} \right\|_1 \le \epsilon_n , \quad \text{as} \quad n \to 0, \quad \epsilon_n \to 0 . \tag{G1}$$

If the pure state ψ is not an eigenstate of the system Hamiltonian H_2 , i.e. $V_{H_2}(\psi) > 0$, and the rate R > 0, then the purity of coherence of ρ should be infinite, i.e. $P_{H_1}(\rho) = \infty$.

As we saw in lemma 26, for a bounded Hamiltonian H_1 , the purity of coherence $P_{H_1}(\rho) = \infty$, if and only if $[\Pi_{\rho}, H_1] \neq 0$, where Π_{ρ} is the projector to the support of ρ . Therefore, we conclude that if $[\Pi_{\rho}, H_1] = 0$, then distillable coherence is zero. This is the case, in particular, for a typical mixed state, which has a full-rank density operator.

Proof. In the following, to simplify the notation, we assume the Hamiltonians of the input and output systems are identical, and both are denoted by H. Generalization to the general case with different Hamiltonians at the input and output is straightforward. Also, without loss of the generality, we assume the average energy of state $|\psi\rangle$ is zero, i.e. $\langle\psi|H|\psi\rangle=0$ (which can always be achieved by a proper choice of energy reference).

The main idea is that if $P_H(\rho) < \infty$ then the purity of coherence of the input will grow linearly with n, the number of copies, as $n \times P_H(\rho)$. On the other hand, if R > 0 and $V_H(\psi) > 0$, then for the output to converge to state $\psi^{\otimes \lceil Rn \rceil}$, the purity of coherence should grow faster than linear in n.

To find a lower bound on the purity of coherence of the output, we use corollary 29 which bounds the purity of coherence for mixed states close to pure states. Let $\sigma = \mathcal{E}_n(\rho^{\otimes n})$ be the actual output state. By assumption, $\|\psi^{\otimes \lceil Rn \rceil} - \sigma\|_1 \leq \epsilon_n$. Then, according to the corollary 29, there exists a pure state $|\Theta_n\rangle$ (namely the eigenvector of σ with the largest eigenvalue) whose overlap with $\psi^{\otimes \lceil Rn \rceil}$ is lower bounded by $|\langle \Theta_n | \psi \rangle^{\otimes \lceil Rn \rceil}|^2 \geq 1 - \epsilon_n$, such that

$$P_{H_{\text{tot}}}(\sigma) \ge V_{H_{\text{tot}}}(|\Theta_n\rangle) \times (2\epsilon_n^{-1} - 3)$$
, (G2)

where $H_{\mathrm{tot}} = \sum_{i=1}^{\lceil Rn \rceil} H^{(i)}$ is the sum of the Hamiltonians of the output systems.

On the other hand, because \mathcal{E}_n is a TI operation, the purity of coherence of the output is upper bounded by the purity of coherence of the input, which by the additivity of the purity of coherence, is $n \times P_H(\rho)$. Therefore, $P_{H_{tot}}(\sigma) \leq nP_H(\rho)$. Combining these two bounds we arrive at

$$nP_H(\rho) \ge V_{H_{rel}}(|\Theta_n\rangle) \times (2\epsilon_n^{-1} - 3)$$
, (G3)

or equivalently,

$$\frac{\epsilon_n}{2 - 3\epsilon_n} \times P_H(\rho) \ge \frac{1}{n} V_{H_{\text{tot}}}(|\Theta_n\rangle) , \tag{G4}$$

where $|\Theta_n\rangle$, is the eigenvector of the output state σ with the largest eigenvalue, and satisfies $|\langle\Theta_n|\psi\rangle^{\otimes \lceil Rn\rceil}|^2 \geq 1 - \epsilon_n$. Next, note that if $P_H(\rho) < \infty$, then the left-hand side vanishes in the limit n goes to infinity, which implies

$$\lim_{n \to \infty} \frac{1}{n} V_{H_{\text{tot}}}(|\Theta_n\rangle) = 0.$$
 (G5)

In the following, we argue that this if R>0 and $V_H(\psi)>0$, then this equation is in contradiction with $\lim_{n\to\infty}|\langle\Theta_n|\psi\rangle^{\otimes \lceil Rn\rceil}|^2=1$. In other words, state $|\Theta_n\rangle$ does not have enough energy variance (Quantum Fisher Information) to converge to state $|\psi\rangle^{\otimes \lceil Rn\rceil}$. To prove this we use the central limit theorem.

In the limit $n \to \infty$, the energy distribution for state $|\psi\rangle^{\otimes \lceil Rn \rceil}$ converges to a Gaussian distribution with mean zero (Recall that we assume $\langle \psi | H | \psi \rangle = 0$). More precisely, define the observable

$$\tilde{H}_{\text{tot}} = \frac{1}{\sqrt{\lceil Rn \rceil}} H_{\text{tot}} = \frac{1}{\sqrt{\lceil Rn \rceil}} \sum_{i=1}^{\lceil Rn \rceil} H^{(i)} . \tag{G6}$$

Then, by the central limit theorem, the distribution over eigenvalues of \tilde{H}_{tot} , for state $|\psi\rangle^{\otimes \lceil Rn \rceil}$ converges to a Gaussian with mean zero, and variance

$$V_{\tilde{H}_{H}}(\psi^{\otimes \lceil Rn \rceil}) = V_H(\psi) > 0. \tag{G7}$$

In particular, this means that in the limit $n \to \infty$ there is a non-vanishing probability that state $|\psi\rangle^{\otimes \lceil Rn \rceil}$ can be found in the subspaces corresponding to the eigenvalues λ of observable \tilde{H}_{tot} with $\lambda > \sqrt{V_H(\psi)}$ (or with $\lambda < -\sqrt{V_H(\psi)}$). In other words, relative to the total Hamiltonian $H_{\text{tot}} = \sum_{i=1}^{\lceil Rn \rceil} H^{(i)}$, there is a non-zero (order one) probability that state $\psi^{\otimes \lceil Rn \rceil}$ is found in the energies higher than $\sqrt{n \times RV_H(\psi)}$ or lower than $-\sqrt{n \times RV_H(\psi)}$.

On the other hand, for state $|\Theta_n\rangle$, we have $\lim_{n\to\infty}\frac{1}{n}V_{H_{tot}}(|\Theta_n\rangle)=0$. Using the fact that

$$\frac{1}{n}V_{H_{\text{tot}}}(|\Theta_n\rangle) = \frac{\lceil Rn \rceil}{n} \times V_{\tilde{H}_{\text{tot}}}(|\Theta_n\rangle) , \qquad (G8)$$

this implies

$$\lim_{n \to \infty} V_{\tilde{H}_{\text{tot}}}(|\Theta_n\rangle) = 0 , \qquad (G9)$$

i.e. in the limit $n \to \infty$ the distribution over eigenvalues of \tilde{H}_{tot} corresponding to state $|\Theta_n\rangle$ has variance zero. Equivalently, this means that in the limit $n \to \infty$, the energy distribution of state $|\Theta_n\rangle$ (relative to the total Hamiltonian $H_{\text{tot}} = \sum_{i=1}^{\lceil Rn \rceil} H^{(i)}$) cannot have a non-vanishing probability in both regions $E > \sqrt{n \times RV_H(\psi)}$ and $E < -\sqrt{n \times RV_H(\psi)}$.

It follows that, even in the limit $n \to \infty$, there is a non-zero (order one) probability of distinguishing the distributions over eigenvalues of \tilde{H}_{tot} (or, equivalently, H_{tot}) for the two states $|\Theta_n\rangle$ and $|\psi\rangle^{\otimes \lceil Rn \rceil}$. But, this is in contradiction with $\lim_{n\to\infty} |\langle \Theta_n|\psi\rangle^{\otimes \lceil Rn \rceil}|^2=1$, because the latter implies that the probability of distinguishing the two states via any kind of measurement should go to zero, in the limit $n\to\infty$. Therefore, we conclude that if $P_H(\rho)<\infty$, then R=0, or $V_H(\psi)=0$.

It turns out that this result can be extended to the case where one is allowed to use a finite *helper* system at the input to implement the transformation

$$\rho^{\otimes n} \otimes \chi \xrightarrow{TI} \stackrel{\epsilon_n}{\approx} \psi^{\otimes \lceil Rn \rceil}, \tag{G10}$$

where χ is the state of the helper system, in a finite-dimensional Hilbert space with a bounded Hamiltonian. The helper system can be in a pure state, in which case the purity of coherence of the input can be ∞ , even for finite n. Therefore, in this case it is not clear that how we can put a restriction on the output based on the purity of coherence of the input. Nevertheless, we can overcome this issue, and prove an extension of theorem 30, which implies distillable coherence remains zero for states with bounded purity of coherence, even if one allows a finite helper system at the input.

This extension of theorem 30, follows from the following lemma together with the argument we used to prove theorem 30.

Lemma 31. Suppose there exists a TI operation \mathcal{E}_n which transforms n copies of system with state ρ and Hamiltonian H and a helper system in state χ and Hamiltonian H_{help} , to m copies of a system with Hamiltonian H and state ψ with error ϵ_n in trace distance, such that

$$\left\| \mathcal{E}_n(\rho^{\otimes n} \otimes \chi) - \psi^{\otimes m} \right\|_1 \le \epsilon_n \ . \tag{G11}$$

Then, there exists a pure state $|\Theta_n\rangle$ (namely the eigenstate of $\mathcal{E}_n(\rho^{\otimes n}\otimes\chi)$ with the largest eigenvalue) whose overlap with the desired state $\psi^{\otimes m}$ is

$$|\langle \Theta_n | \psi \rangle^{\otimes m}|^2 \ge 1 - 2\epsilon_n,\tag{G12}$$

and satisfies

$$\epsilon_n P_H(\rho) + 2(d_{\chi} - 1) \frac{1}{n} V_{H_{help}}(\chi) \ge \frac{1}{n} V_{H_{tot}}(|\Theta_n\rangle) \times (1 - 3\epsilon_n), \tag{G13}$$

where d_{χ} is the dimension of the Hilbert space of the helper system, and $H_{tot} = \sum_{i=1}^{m} H^{(i)}$ is the sum of the Hamiltonians of the output systems.

Proof. In general, at the presence of the helper state, the purity of coherence of the input can be ∞ for a finite n, in which case

we cannot put any constraint on the output based on its purity of coherence. To rectify this issue we use the following trick, which can be used more generally when one deals with the purity of coherence for pure states: assume instead of using the helper state in the pure state χ , we use τ_{χ} , a noisy version of χ obtained by mixing χ with the totally mixed state, with a ratio such that the trace distance between χ and τ_{χ} is exactly ϵ_n . Now suppose in the process $\rho^{\otimes n} \otimes \chi \xrightarrow{\mathrm{TI}} \psi^{\otimes m}$, we use τ_{χ} instead of χ . Then, we introduce an additional error in the process. Using the fact that the trace distance satisfies the triangle inequality, and is non-increasing under CPTP maps, this additional error can be bounded by ϵ_n . Therefore, the total error at the output will be bounded by $2\epsilon_n$. To summarize, if

$$\left\| \mathcal{E}_n(\rho^{\otimes n} \otimes \chi) - \psi^{\otimes m} \right\|_1 \le \epsilon_n , \tag{G14}$$

then,

$$\left\| \mathcal{E}_n(\rho^{\otimes n} \otimes \tau_{\chi}) - \psi^{\otimes m} \right\|_1 \le 2\epsilon_n \ . \tag{G15}$$

In this transformation, the purity of coherence for the input is $nP_H(\rho) + P_{H_{\text{help}}}(\tau_\chi)$, which is bounded. Later, we show that $P_{H_{\text{help}}}(\tau_\chi)$ is upper bounded by

$$P_{H_{\text{help}}}(\tau_{\chi}) \le \frac{2(d_{\chi} - 1)}{\epsilon_n} V_{H_{\text{help}}}(\chi) , \qquad (G16)$$

where d_{χ} is the dimension of the Hilbert space of χ . Therefore, the total purity of coherence for the input $\rho^{\otimes n} \otimes \tau_{\chi}$ is lower bounded by

$$nP_H(\rho) + \frac{2(d_{\chi} - 1)}{\epsilon_n} V_{H_{\text{help}}}(\chi) . \tag{G17}$$

Next, we focus on the purity of coherence of the output, and use corollary 29 which bounds the purity of coherence for mixed states close to pure states. Let $\sigma = \mathcal{E}_n(\rho^{\otimes n} \otimes \tau_\chi)$ be the actual output state. By assumption, $\|\psi^{\otimes m} - \sigma\|_1 \leq 2\epsilon_n$. Then, according to the corollary 29, there exists a pure state $|\Theta_n\rangle$ (namely the eigenvector of σ with the largest eigenvalue) whose overlap with $\psi^{\otimes m}$ is lower bounded by

$$\left| \langle \Theta_n | \psi \rangle^{\otimes m} \right|^2 \ge 1 - 2\epsilon_n \,, \tag{G18}$$

such that

$$P_{H_{\text{tot}}}(\sigma) \ge V_{H_{\text{tot}}}(|\Theta_n\rangle) \times (\epsilon_n^{-1} - 3) , \qquad (G19)$$

where $H_{\mathrm{tot}} = \sum_{i=1}^{m} H^{(i)}$ is the sum of the Hamiltonians of the output systems.

Therefore, using the monotonicity of the purity of coherence, we conclude

$$V_{H_{\text{tot}}}(|\Theta_n\rangle) \times (\epsilon_n^{-1} - 3) \le nP_H(\rho) + \frac{2(d_{\chi} - 1)}{\epsilon_n} V_{H_{\text{help}}}(\chi) , \qquad (G20)$$

or equivalently

$$\epsilon_n P_H(\rho) + 2(d_{\chi} - 1) \frac{1}{n} V_{H_{\text{help}}}(\chi) \ge \frac{1}{n} V_{H_{\text{tot}}}(|\Theta_n\rangle) \times (1 - 3\epsilon_n). \tag{G21}$$

To complete the proof, in the following we prove Eq.(G16): Let τ_{χ} be the state obtained by mixing the pure state χ and the totally mixed state I/d_{χ} , such that the trace distance between τ_{χ} and χ is ϵ . Then,

$$\tau_{\chi} = (1 - \frac{\epsilon}{2})|\chi\rangle\langle\chi| + \frac{\epsilon}{2(d_{\chi} - 1)}(I - |\chi\rangle\langle\chi|). \tag{G22}$$

Recall that for any Hamiltonian H and state ρ with spectral decomposition $\rho = \sum_j p_j |\psi_j\rangle \langle \psi_j|$, we have $P_H(\rho) = \sum_{k,l} \frac{p_k^2 - p_l^2}{p_l} |\langle \psi_k | H |\psi_l \rangle|^2$. Therefore, for any j and k whose corresponding eigenvalues are equal the corresponding term

 $\frac{p_k^2-p_l^2}{p_l}|\langle\psi_k|H|\psi_l\rangle|^2$ does not contribute in the summation. Using this for state au_χ , we find

$$P_{H_{\text{help}}}(\tau_{\chi}) \le \frac{(1 - \frac{\epsilon}{2})^2 - (\frac{\epsilon}{2(d_{\chi} - 1)})^2}{\frac{\epsilon}{2(d_{\chi} - 1)}} \sum_{l:\psi_l \ne \chi} |\langle \chi | H_{\text{help}} | \psi_l \rangle|^2$$
(G23)

$$\leq \frac{2(d_{\chi} - 1)}{\epsilon} V_{H_{\text{help}}}(\chi) . \tag{G24}$$

This completes the proof of the lemma.

Appendix H: Single-copy distillation

1. Maximum achievable fidelity with a pure state

In this section, using the results of [79], we present a simple formula for the maximum achievable fidelity $\max_{\mathcal{E}_{\text{TI}}} \langle \psi | \mathcal{E}_{\text{TI}}(\rho^{\otimes n}) | \psi \rangle$, in terms of the *conditional min-entropy*.

Recall the definition of the conditional min-entropy, $H_{\min}(B|A)_{\Omega}$, of a bipartite state Ω^{AB} ,

$$2^{-H_{\min}(B|A)_{\Omega}} = \inf_{\tau^A \ge 0} \left\{ \operatorname{Tr}(\tau^A) : \tau^A \otimes I^B \ge \Omega^{AB} \right\}. \tag{H1}$$

The following result follows from the arguments of [79].

Proposition 32. [79] Let H_A and H_B be, respectively, the Hamiltonians of the input and output systems A and B. Let σ_A and be an arbitrary state of A and $|\psi\rangle_B$ be a pure state of system B. Then,

$$\max_{\mathcal{E}_{TI}} \langle \psi | \mathcal{E}_{TI}(\sigma_A) | \psi \rangle_B = 2^{-H_{min}(B|A)_{\Omega}} , \tag{H2}$$

where the maximization is over the set of all TI operations, and state Ω_{AB} is defined as

$$\Omega_{AB} = \lim_{T \to \infty} \frac{1}{T} \int_0^T dt \ e^{-iH_A t} \otimes e^{iH_B t} [\sigma_A \otimes |\overline{\psi}\rangle \langle \overline{\psi}|_B] e^{-iH_A t} \otimes e^{iH_B t}$$
(H3)

$$= \sum_{E} \Pi_{E} [\sigma_{A} \otimes |\overline{\psi}\rangle \langle \overline{\psi}|_{B}] \Pi_{E} , \qquad (H4)$$

where $|\overline{\psi}\rangle_B$ is the complex conjugate of $|\psi\rangle_B$ in the energy eigenbasis, and Π_E is the projector to the eigensubspace of $H_A \otimes I_B - I_A \otimes H_B$ with energy E.

In other words, state Ω_{AB} is the state obtained by dephasing $\sigma_A \otimes |\overline{\psi}\rangle \langle \overline{\psi}|_B$ in the eingenbasis of the Hamiltonian $H_A \otimes I_B - I_A \otimes H_B$.

Note that if the input system A is n copies of a system with Hamiltonian H and state ρ , then state Ω_{AB} will be given by

$$\sum_{E} \Pi_{E}[\rho^{\otimes n} \otimes |\overline{\psi}\rangle \langle \overline{\psi}|_{B}] \Pi_{E} , \qquad (H5)$$

where Π_E is the projector to the eigensubpaces of Hamiltonian $H_{\text{tot}} \otimes I_B - I_{\text{tot}} \otimes H_B$, where $H_{\text{tot}} = \sum_i H^{(i)}$ is the total Hamiltonian of the input systems, $H^{(i)} = I^{\otimes (i-1)} \otimes H \otimes I^{\otimes (n-i-1)}$, and $I_{\text{tot}} = I^{\otimes n}$ is the identity operator on the input systems.

Remark 33. This result can be easily extended to the case of other symmetries with finite or compact Lie groups. In this case state Ω_{AB} will be defined by averaging over the Haar measure of the group under consideration.

Proof. Let B' be an auxiliary system with dimension equal to d_B , the dimension B. Let $\{|E_i\rangle_B: i=1,\cdots,d_B\}$ be the eigenstates of Hamiltonian H_B , and

$$|\gamma_{B'B}\rangle = \frac{1}{\sqrt{d_B}} \sum_{i=1}^{d_B} |E_i E_i\rangle_{B'B} \tag{H6}$$

be a maximally entangled state of B and the auxiliary system B'. Then, for any pair of operators X and Y defined on B, we have $\text{Tr}(XY) = d_B \times \langle \gamma_{B'B} | [X \otimes Y^T] | \gamma_{B'B} \rangle$, where T denotes transpose in the energy eigenbasis, $\{|E_i\rangle_B : i = 1, \cdots, d_B\}$.

This implies that for any quantum channel \mathcal{E}_{TI} we have

$$\langle \psi | \mathcal{E}_{\text{TI}}(\sigma) | \psi \rangle = d_B \times \langle \gamma_{B'B} | \left[\mathcal{E}_{\text{TI}}(\sigma) \otimes | \overline{\psi} \rangle \langle \overline{\psi} | \right] | \gamma_{B'B} \rangle , \tag{H7}$$

where $|\overline{\psi}\rangle$ is the complex conjugate of $|\psi\rangle$ in the energy eigenbasis.

Next, we note that

$$\langle \psi | \mathcal{E}_{\text{TI}}(\sigma) | \psi \rangle = d_B \times \langle \gamma_{B'B} | \left[\mathcal{E}_{\text{TI}}(\sigma) \otimes | \overline{\psi} \rangle \langle \overline{\psi} | \right] | \gamma_{B'B} \rangle \tag{H8}$$

$$= d_B \times \langle \gamma_{B'B} | (e^{iH_B t} \otimes e^{-iH_B t}) \left[\mathcal{E}_{\text{TI}}(\sigma) \otimes |\overline{\psi}\rangle \langle \overline{\psi}| \right] (e^{-iH_B t} \otimes e^{iH_B t}) |\gamma_{B'B}\rangle$$
(H9)

$$= d_B \times \langle \gamma_{B'B} | (\mathcal{E}_{\text{TI}} \otimes \mathcal{I}) \left([e^{iH_A t} \otimes e^{-iH_B t}] [\sigma \otimes |\overline{\psi}\rangle \langle \overline{\psi}|] [e^{-iH_A t} \otimes e^{iH_B t}] \right) |\gamma_{B'B}\rangle , \tag{H10}$$

where to get the second line we have used the fact that

$$(e^{-iH_Bt} \otimes e^{iH_Bt})|\gamma_{B'B}\rangle = |\gamma_{B'B}\rangle, \tag{H11}$$

and to get the last line we have used the fact that \mathcal{E}_{TI} satisfies the covariance condition

$$\mathcal{E}_{\text{TI}}(e^{-iH_At}(\cdot)e^{iH_At}) = e^{-iH_Bt}\mathcal{E}_{\text{TI}}(\cdot)e^{iH_Bt}, \ \forall t \in \mathbb{R}.$$
(H12)

Then, taking the average over t, we find that

$$\langle \psi | \mathcal{E}_{\text{TI}}(\sigma) | \psi \rangle = d_B \times \langle \gamma_{B'B} | \mathcal{E}_{\text{TI}} \otimes \mathcal{I}(\Omega_{AB}) | \gamma_{B'B} \rangle , \qquad (\text{H}13)$$

where

$$\Omega_{AB} \equiv \lim_{T \to \infty} \frac{1}{T} \int_0^T dt \, \left([e^{iH_A t} \otimes e^{-iH_B t}] [\sigma \otimes |\overline{\psi}\rangle \langle \overline{\psi}|] [e^{-iH_A t} \otimes e^{iH_B t}] \right) \,. \tag{H14}$$

Therefore,

$$\max_{\mathcal{E}_{\text{TI}}} \langle \psi | \mathcal{E}_{\text{TI}}(\sigma) | \psi \rangle = d_B \times \max_{\mathcal{E}_{\text{TI}}} \langle \gamma_{B'B} | \mathcal{E}_{\text{TI}} \otimes \mathcal{I}(\Omega_{AB}) | \gamma_{B'B} \rangle . \tag{H15}$$

On the other hand, using the covariance condition in Eq.(H12) and symmetry of state $|\gamma_{B'B}\rangle$ in Eq.(H11) we can easily see that for any general quantum operation \mathcal{E} , there exists a TI operation \mathcal{E}_{TI} , defined by

$$\mathcal{E}_{\text{TI}}(X) \equiv \lim_{T \to \infty} \frac{1}{T} \int_0^T dt \ e^{iH_B t} \mathcal{E}(e^{-iH_A t}(X)e^{iH_A t})e^{-iH_B t} , \qquad (\text{H}16)$$

such that

$$\langle \gamma_{B'B} | \mathcal{E} \otimes \mathcal{I}(\Omega_{AB}) | \gamma_{B'B} \rangle = \langle \gamma_{B'B} | \mathcal{E}_{TI} \otimes \mathcal{I}(\Omega_{AB}) | \gamma_{B'B} \rangle. \tag{H17}$$

This implies that in the right-hand side of Eq.(H15), maximization over all TI quantum operations, can be replaced by maximization over all quantum operations, i.e.

$$\max_{\mathcal{E}_{\Pi}} \langle \psi | \mathcal{E}_{\Pi}(\sigma) | \psi \rangle = d_B \times \max_{\mathcal{E}_{\Pi}} \langle \gamma_{B'B} | \mathcal{E}_{\Pi} \otimes \mathcal{I}(\Omega_{AB}) | \gamma_{B'B} \rangle$$
(H18)

$$= d_B \times \max_{\mathcal{E}} \langle \gamma_{B'B} | \mathcal{E} \otimes \mathcal{I}(\Omega_{AB}) | \gamma_{B'B} \rangle. \tag{H19}$$

Finally, using the result of [82] (See also [78]) we note that

$$2^{-H_{\min}(B|A)_{\Omega}} = d_B \times \max_{\mathcal{E}} \langle \gamma_{B'B} | \mathcal{E} \otimes \mathcal{I}_B(\Omega_{AB}) | \gamma_{B'B} \rangle , \qquad (\text{H20})$$

where the maximization if over all CPTP maps from system A to system B'. Therefore, we conclude that

$$2^{-H_{\min}(B|A)_{\Omega}} = \max_{\mathcal{E}_{\Pi}} \langle \psi | \mathcal{E}_{\Pi}(\sigma) | \psi \rangle . \tag{H21}$$

2. Qubit example

The smallest quantum clock is a qubit with two different energy levels. For a two-level system with Hamiltonian $H = \pi \sigma_z / \tau$, suppose we want to prepare a qubit clock in a state close to the pure state $|\Phi\rangle_{\text{c-bit}} = (|0\rangle + |1\rangle)/\sqrt{2}$. Assume we start with n copies of a noisy version of state $|\Phi\rangle_{\text{c-bit}}$, i.e. state

$$\rho = \lambda |\Phi\rangle\langle\Phi|_{\text{c-bit}} + (1-\lambda)I/2, \qquad (H22)$$

with $0 < \lambda < 1$, and we want to purify this state via TI operations and obtain a qubit state σ which has higher fidelity with $|\Phi\rangle_{\text{c-bit}}$. How close can we get to state $|\Phi\rangle_{\text{c-bit}}$? In other words, what is the maximum achievable fidelity,

$$\max_{\mathcal{E}_{\text{TI}}} \langle \Phi | \mathcal{E}_{\text{TI}}(\rho^{\otimes n}) | \Phi \rangle_{\text{c-bit}} , \qquad (\text{H23})$$

where the maximization is over all TI operations.

For any TI operation \mathcal{E}_{TI} , let $\sigma = \mathcal{E}_{TI}(\rho^{\otimes n})$ be the actual output state of the transformation. Then, using the monotonicity and the additivity of the purity of coherence we have

$$P_H(\sigma) \le P_H(\rho^{\otimes n}) = nP_H(\rho) . \tag{H24}$$

As we saw in Eq.(F42), for a general qubit state ρ with the spectral decomposition $\gamma = p|\psi\rangle\langle\psi| + (1-p)|\psi^{\perp}\rangle\langle\psi^{\perp}|$, the purity of coherence is given by

$$P_H(\gamma) = \frac{(1-2p)^2}{p(1-p)} \times V_H(\psi) . \tag{H25}$$

For state $\rho = \lambda |\Phi\rangle\langle\Phi|_{\text{c-bit}} + (1-\lambda)I/2$, we have $p = (1+\lambda)/2$, and $\psi = \Phi_{\text{c-bit}}$. Therefore,

$$P_H(\rho) = \frac{4\lambda^2}{1 - \lambda^2} \times V_H(\Phi_{\text{c-bit}}) . \tag{H26}$$

We conclude that for the output state σ , it holds that

$$P_H(\sigma) \le \frac{4n\lambda^2}{1-\lambda^2} \times V_H(\Phi_{\text{c-bit}}) . \tag{H27}$$

Using the symmetries of this problem, it can be easily seen that the optimal fidelity can be achieved for an output state in the form of

$$\sigma = \tilde{\lambda} |\Phi\rangle \langle \Phi|_{\text{c-bit}} + (1 - \tilde{\lambda})I/2 , \qquad (H28)$$

for some $0 < \tilde{\lambda} \le 1$. For this state,

$$P_H(\sigma) = \frac{4\tilde{\lambda}^2}{1 - \tilde{\lambda}^2} \times V_H(\Phi_{\text{c-bit}}) . \tag{H29}$$

Putting this into Eq.(H27) we find

$$\frac{\tilde{\lambda}^2}{1 - \tilde{\lambda}^2} \le n \times \frac{\lambda^2}{1 - \lambda^2} \,. \tag{H30}$$

which implies

$$\tilde{\lambda}^2 \le \frac{n\lambda^2}{1 + (n-1)\lambda^2} = \frac{1}{1 + \frac{1}{\pi}(\frac{1}{\lambda^2} - 1)} \,. \tag{H31}$$

For a fixed $\lambda > 0$, in the large n limit this implies

$$\tilde{\lambda}^2 \le \frac{1}{1 + \frac{1}{n}(\frac{1}{\lambda^2} - 1)} = 1 - \frac{1}{n}(\frac{1 - \lambda^2}{\lambda^2}) + \mathcal{O}(\frac{1}{n^2}). \tag{H32}$$

Therefore,

$$1 - \tilde{\lambda}^2 \ge \frac{1}{n} \frac{1 - \lambda^2}{\lambda^2} + \mathcal{O}(\frac{1}{n^2}). \tag{H33}$$

Finally, we note that $\mathcal{E}_{\mathrm{TI}}(\rho^{\otimes n})=\sigma=\tilde{\lambda}|\Phi\rangle\langle\Phi|_{\mathrm{c-bit}}+(1-\tilde{\lambda})I/2$ implies that

$$1 - \langle \Phi | \mathcal{E}_{\text{TI}}(\rho^{\otimes n}) | \Phi \rangle_{\text{c-bit}} = 1 - [\tilde{\lambda} + (1 - \tilde{\lambda})/2] = 1 - (1 + \tilde{\lambda})/2 = (1 - \tilde{\lambda})/2 . \tag{H34}$$

Given that $\tilde{\lambda} \leq 1$, we find $1 \geq (1 + \tilde{\lambda})/2$, which implies

$$1 - \langle \Phi | \mathcal{E}_{\text{TI}}(\rho^{\otimes n}) | \Phi \rangle_{\text{c-bit}} = (1 - \tilde{\lambda})/2 \tag{H35}$$

$$\geq (1 - \tilde{\lambda}^2)/4 \tag{H36}$$

$$\geq \frac{1}{4} \frac{1}{n} \frac{1 - \lambda^2}{\lambda^2} + \mathcal{O}(\frac{1}{n^2}). \tag{H37}$$

Interestingly, using the result of Cirac et al [84], we know that there exists a TI operation which achieves

$$1 - \langle \Phi | \mathcal{E}_{\text{TI}}(\rho^{\otimes n}) | \Phi \rangle_{\text{c-bit}} = \frac{1}{2} \frac{1}{n} \frac{1 - \lambda}{\lambda^2} + \mathcal{O}(\frac{1}{n^2}) . \tag{H38}$$