Thermodynamic Capacity of Quantum Processes

Philippe Faist,[1,*] Mario Berta,[2] and Fernando Brandão[3,1]

[1] *Institute for Quantum Information and Matter, Caltech, Pasadena, CA 91125, U.S.A.*
[2] *Department of Computing, Imperial College London, London SW7 2AZ, United Kingdom*
[3] *Google Inc., Venice, CA 90291, U.S.A.*

(Dated: July 12, 2018)

Thermodynamics imposes restrictions on what state transformations are possible. In the macroscopic limit of asymptotically many independent copies of a state—as for instance in the case of an ideal gas—the possible transformations become reversible and are fully characterized by the free energy. In this letter, we present a thermodynamic resource theory for quantum processes that also becomes reversible in the macroscopic limit. Namely, we identify a unique single-letter and additive quantity, the thermodynamic capacity, that characterizes the "thermodynamic value" of a quantum channel. As a consequence the work required to simulate many repetitions of a quantum process employing many repetitions of another quantum process becomes equal to the difference of the respective thermodynamic capacities. For our proof, we construct explicit universal implementations of quantum processes using Gibbs-preserving maps and a battery, requiring an amount of work asymptotically equal to the thermodynamic capacity. This implementation is also possible with thermal operations in the case of time-covariant quantum processes or when restricting to independent and identical inputs. In our derivations we make extensive use of Schur-Weyl duality and other information-theoretic tools, leading to a generalized notion of quantum typical subspaces.

Introduction.—In the quest of extending the laws of thermodynamics beyond the macroscopic regime, the resource theory of thermal operations was introduced to characterize possible transformations which could be carried out at the nano scale [1–5]. By imposing a set of physically motivated rules, an agent can only perform a restricted set of operations on a system, which we refer to generically as *thermodynamic operations*. By characterizing the possible state transformations under these rules, one obtains formulations of the second law of thermodynamics which are valid for small-scale systems out of thermodynamic equilibrium. A natural regime to study such state transformations is a macroscopic regime in which one considers conversions between many independent and identically distributed (i.i.d.) copies of a state, i.e., states of the form $\rho^{\otimes n}$. If we consider transformations on a system S with Hamiltonian H_S, using a heat bath at inverse temperature β and a work storage system W, then the asymptotic work cost per copy of transforming $\rho^{\otimes n}$ into $\sigma^{\otimes n}$ is given by the difference in free energy $F(\sigma) - F(\rho)$. The free energy is defined as

$$F(\rho) = \mathrm{tr}(H\rho) - \beta^{-1}S(\rho) = \beta^{-1}D\big(\rho\,\big\|\,e^{-\beta H}\big)\,, \qquad (1)$$

expressed either in terms the von Neumann entropy $S(\rho) := -\mathrm{tr}[\rho \ln \rho]$ and the quantum relative entropy $D(\rho\,\|\,\gamma) := \mathrm{tr}[\rho(\ln\rho - \ln\gamma)]$. Since the cost of asymptotically performing the reverse transformation $\sigma^{\otimes n} \to \rho^{\otimes n}$ is the negative of the cost of the forward transformation this resource theory becomes reversible (Fig. 1a).

Here, we study the resource theory of thermodynamics for quantum processes themselves. Given a black-box implementation of a process $\mathcal{E}$, can we simulate a process

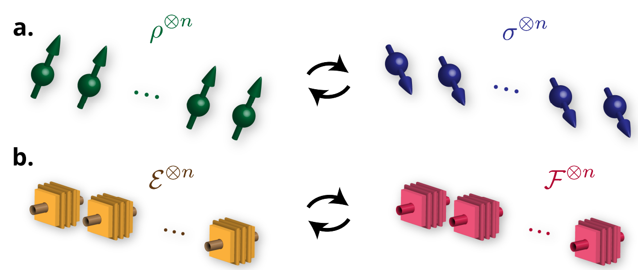

FIG. 1: Asymptotic reversible interconversion of quantum channels. **a.** In the resource theory of thermodynamics, quantum states are reversibly inter-convertible, i.e., the work cost of transforming n independent copies of ρ into n independent copies of σ is (approximately) the same as the work that can be extracted in the reverse transformation. **b.** We show that a similar conclusion can be drawn for quantum processes. There is a unique quantity, the thermodynamic capacity, that measures the "thermodynamic value" of the channel in terms of resources required to create, or extracted while consuming, many copies of a channel.

$\mathcal{F}$ using thermodynamic operations, or is there a thermodynamic cost in doing so? We fully answer the question in the i.i.d. regime, and show that the thermodynamic simulation of channels becomes reversible (Fig. 1b). That is, the work cost of executing many realizations of $\mathcal{F}$ using many realizations of $\mathcal{E}$ is the same as the work that can be extracted in the reverse process of implementing $\mathcal{E}$ from uses of $\mathcal{F}$.

The underlying mathematical problem which we solve is that of implementing a given quantum process accurately for any possible input state using thermodynamic operations (Fig. 2). As a starting point, in the case where we look at one repetition of the quantum process and the

* phfaist@caltech.edu

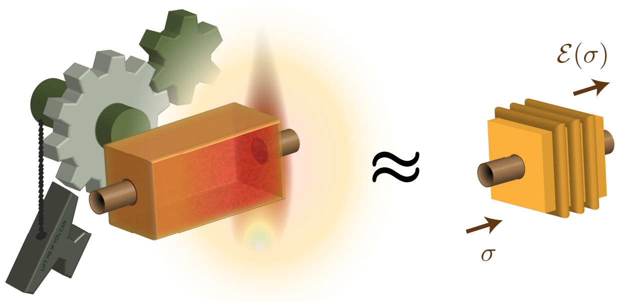

FIG. 2: Universal thermodynamic implementation of a quantum process. Using thermodynamic operations and by furnishing work, the task is to simulate an ideal process $\mathcal{E}$, up to an approximation error measured in diamond norm, meaning that the implementation outputs a state close to $\mathcal{E}(\sigma)$ for any possible input σ (even relative to a reference system). In this letter, we consider the regime of many independent copies of the channel, the i.i.d. regime, and we consider two common frameworks of thermodynamic operations. If we allow any Gibbs-preserving map for free, along with the possibility to lift or to lower a weight, then we show that many copies of any given process $\mathcal{E}$ can be implemented at a work cost that converges asymptotically over many copies to the thermodynamic capacity of the process. The same holds for time-covariant processes in the physical framework of thermal operations, i.e., energy-conserving interactions with a heat bath.

input state is fixed, the work cost is given by the *coherent relative entropy* [6]. The coherent relative entropy is a one-shot information measure that generalizes the quantum conditional entropy as well as the quantum relative entropy. We note that thermodynamic tasks are often closely related to one-shot information measures [2, 3, 7–9], and using known tools from one-shot information theory it follows that the work cost rate of implementing an i.i.d. channel $\mathcal{E}^{\otimes n}$ for a fixed i.i.d. input state $\sigma^{\otimes n}$ becomes asymptotically equal to $F(\mathcal{E}(\sigma)) - F(\sigma)$. Our main technical contribution is to construct an optimal thermodynamic implementation of an i.i.d. quantum process, that is universal, i.e., that implements the process accurately for any input state. Crucially, the implementation does not depend on the input state, and when considered as a channel, is close in diamond norm to the ideal channel $\mathcal{E}^{\otimes n}$. As we will see the rate at which work has to be supplied is characterized by the *thermodynamic capacity* of the channel, given as

$$T(\mathcal{E}) = \max_{\sigma} \left\{ F(\mathcal{E}(\sigma)) - F(\sigma) \right\}. \tag{2}$$

That is, the work cost of such an implementation coincides with the worst-case cost of implementing the process for the possible i.i.d. input states. The thermodynamic capacity generalizes the notion of capacity for quantum channels to the context of thermodynamics, by measuring "how much free energy" can be conveyed through the use of the channel. Combining our main result with the fact that it is possible to extract an amount of work equal to $T(\mathcal{E})$ from a black-box implementation of $\mathcal{E}$ [10], we find that the thermodynamic capacity is the unique measure of the "thermodynamic value" of the quantum channel. Hence, the corresponding resource theory is reversible. We note that the thermodynamic capacity in (2) is expressed as a single-letter formula and is additive [10].

We prove our main result for two different thermodynamic frameworks. First, for any i.i.d. process we construct a universal implementation using the framework of Refs. [6, 11]. Here, one allows for free any map that preserves the Gibbs state, and counts work using an explicit work storage system. Second, we consider the framework of thermal operations, in which energy-conserving interactions are allowed with a heat bath [1–3], and work is again counted using an explicit storage system. For this case we construct a universal protocol that optimally implements any i.i.d. channel that is time-covariant, i.e., that commutes with the time evolution. Thermal operations are a more physical set of operations to consider when developing thermodynamic protocols compared to Gibbs-preserving maps, as it is not clear under which conditions one can implement the latter.

We also provide a collection of secondary, related results. This includes a conceptually direct proof of the asymptotic equipartition property of the coherent relative entropy as well as a novel procedure for Landauer erasure with quantum side information for a non-interacting system and memory with non-trivial Hamiltonians, using thermal operations and a battery. Finally, we extend the previous work [6] on Gibbs preserving maps by showing how to implement a general i.i.d. process for a fixed i.i.d. input state using only thermal operations and a battery.

Our proofs rely on tools from Schur-Weyl duality [12–14] as well as the related post-selection technique [15]. A main component is a new fully quantum universal smoothing operator for bipartite states that counts entropy relative to another operator, and which is a natural generalization of universal and relative typical subspace projectors [13, 16–24]. For the protocol based on thermal operations, we also adapt information-theoretic techniques based on the convex-split lemma and position-based decoding [25–29].

Thermodynamic simulation of quantum processes.—We first consider the thermodynamic framework of so-called Gibbs-preserving maps, or more generally, Gibbs-sub-preserving maps [6]. In this setting, for a fixed inverse temperature β, one can perform any trace non-increasing, completely positive map Φ_S on a system S with Hamiltonian H_S for free, if it satisfies $\Phi(\Gamma_S) \leqslant \Gamma_S$, where $\Gamma_S = e^{-\beta H_S}$. Such mappings can always be dilated to fully trace preserving maps on a larger system that have the thermal state as a fixed point [6]. Work is accounted for by using an explicit system W, often called *battery*, with a given set of energy levels $\{|E\rangle\}$. For instance, if we can find an allowed thermodynamic operation that transforms $\rho \otimes |0\rangle\langle 0|_W \to \sigma \otimes |E\rangle\langle E|_W$, then we have extracted work E while carrying out the transformation

$\rho \to \sigma$, since our battery is charged by an amount E after the application of a free operation [2, 30, 31]. This work storage model is equivalent to several other common models, such as the information battery [3, 6].

We now study the problem of universally implementing a process using only Gibbs-sub-preserving maps and a battery. More precisely, we require the implementation to be accurate in the *diamond norm distance*, which is defined for two channels $\mathcal{E}, \mathcal{F}$ acting on the same system as

$$\|\mathcal{E} - \mathcal{F}\|_\diamond = \max_\sigma \left\|(\mathcal{E} \otimes \mathrm{id})(\sigma) - (\mathcal{F} \otimes \mathrm{id})(\sigma)\right\|_1 ,$$

where $\|X\|_1 := \mathrm{tr}\big(\sqrt{X^\dagger X}\big)$ denotes the trace norm. In the one-shot regime, we know that any trace non-increasing, completely positive map $\mathcal{T}$ can be implemented exactly with Gibbs-sub-preserving maps, while consuming an amount of work W, if and only if $\mathcal{T}(\Gamma_S) \leqslant e^{\beta W}\,\Gamma_S$ [6]. Hence, the work required to universally implement $\mathcal{E}$ up to a precision ϵ is given by the quantity

$$W^\epsilon(\mathcal{E} \,\|\, \Gamma_S) = \min_{\substack{\mathcal{T}(\Gamma_S) \leqslant e^{\beta w}\,\Gamma_S \\ \frac{1}{2}\|\mathcal{T}-\mathcal{E}\|_\diamond \leqslant \epsilon}} w\,, \tag{3}$$

where the optimization ranges over all completely positive, trace non-increasing maps $\mathcal{T}_{X\to X'}$. Our first main theorem is the following asymptotic equipartition property, which states that there exists a universal thermodynamic implementation of many copies of $\mathcal{E}$ in the framework of Gibbs-sub-preserving maps, using an amount of work per copy that is asymptotically equal to the thermodynamic capacity.

Theorem I. *Let S be a system with Hamiltonian H_S and let $\beta \geqslant 0$. Let $\mathcal{E}$ be any completely positive, trace-preserving map. Then, for any $\epsilon \in (0, 1]$,*

$$\lim_{n\to\infty} \frac{1}{n}\, W^\epsilon(\mathcal{E}^{\otimes n} \,\|\, \Gamma_S^{\otimes n}) = T(\mathcal{E})\,. \tag{4}$$

The full proof is presented in Appendix E. Since the diamond norm distance guarantees an accurate implementation of the channel $\mathcal{E}^{\otimes n}$ for all possible inputs including non-i.i.d. inputs, the proof of Theorem I does not at all follow from standard i.i.d. considerations. We prove Theorem I in two steps, which establish both inequality directions. The easy direction straightforwardly follows from the fact that a universal implementation must in particular be a good implementation for any fixed input state, relating the quantity (3) to the coherent relative entropy whose asymptotics are known. For the hard direction, we first use the post-selection technique to reduce the diamond norm accuracy requirement to a constraint on the trace norm evaluated on a fixed input state known as the de Finetti state [15]. Then, we show that a near-optimal process in the optimization (3) is to first carry out explicitly the purified form of the mappings $\mathcal{E}^{\otimes n}$ using ancillas, and then acting with a smoothing operator similar to a typicality projector. The relevant smoothing operator generalizes the notion of joint typicality projectors [13, 16–24] to fully quantum bipartite states relative to any Γ operators. As such we believe it may be of independent interest.

Proposition II. *Let $\Gamma_{AB}, \Gamma'_B \geqslant 0$ and $x \in \mathbb{R}$. Then, for any $\delta > 0$, there exists $\eta > 0$ and for any $n \in \mathbb{N}$ there exists an operator $M_{A^nB^n}$ satisfying the following properties:*

(i) $M^\dagger_{A^nB^n} M_{A^nB^n} \leqslant \mathbb{1}_{A^nB^n}$;

(ii) $\mathrm{tr}_{A^n}(M\,\Gamma_{AB}^{\otimes n}\,M^\dagger) \leqslant \mathrm{poly}(n)\,e^{-n(x-4\delta)}\,\Gamma_B'^{\otimes n}$;

(iii) We have for any pure state $|\rho\rangle_{ABR}$ satisfying $D(\rho_{AB} \,\|\, \Gamma_{AB}) - D(\rho_B \,\|\, \Gamma'_B) \geqslant x$ that

$$\mathrm{Re}\big\{\langle\rho|_{ABR}^{\otimes n} M_{A^nB^n} |\rho\rangle_{ABR}^{\otimes n}\big\} \geqslant 1 - \mathrm{poly}(n)\,e^{-n\eta}\,,$$

where R denotes a reference system.

As an immediate consequence of Theorem I we show that the thermodynamic resource theory of quantum channels is asymptotically reversible in the i.i.d. regime. We have seen that the thermodynamic capacity quantifies the value of a channel in terms of how much work is needed per copy to universally implement it using thermodynamic operations and a battery. That is, we can "form channels" at that rate. One can also consider the reverse task: In Ref. [10], it is shown that the amount of work that can be extracted per copy from a black-box implementation of the channel is asymptotically also precisely given by the thermodynamic capacity. That is, we can "distill work" at a rate which matches the reverse operation. Hence, any i.i.d. channel transformation $\mathcal{E}^{\otimes n} \to \mathcal{F}^{\otimes n}$ can be carried out by first extracting an amount of work $T(\mathcal{E})$ from each copy of $\mathcal{E}$, and then preparing $\mathcal{F}^{\otimes n}$ at a work cost of $T(\mathcal{F})$ per copy. Clearly, any work invested in the transformation in one direction can be recovered with the reverse transformation.[1] As a result the thermodynamic capacity is the unique monotone that fully characterizes the possible thermodynamic channel conversions, much like the channel's mutual information characterizes the communication value rate of a quantum channel in the presence of entanglement [13, 32].

Corollary III. *Let X be a quantum system with Hamiltonian H_X. Let $\mathcal{E}_X$, $\mathcal{F}_X$ be two completely positive, trace-preserving maps. Then, given a black-box implementation of $\mathcal{E}_X^{\otimes n}$, it is possible to simulate $\mathcal{F}_X^{\otimes n}$ with asymptotically vanishing error using Gibbs-sub-preserving maps and investing an amount of work per copy equal to $T(\mathcal{F}) - T(\mathcal{E})$. This work cost (gain) is optimal.*

[1] Equivalently, we may consider the asymptotic conversion rates $\mathcal{E}^{\otimes n} \to \mathcal{F}^{\otimes rn}$ without investing or extracting work, as is more usually done in the literature of resource theories, and see that the optimal rate is $r = T(\mathcal{E})/T(\mathcal{F})$.

Thermal Operations.—The framework of Gibbs-sub-preserving maps is particularly generous, and it is a priori not clear that all such maps are implementable at no work cost. In the alternative framework of thermal operations, each system S of interest has an associated Hamiltonian H_S and is not interacting with the other systems. For a given fixed inverse temperature β, we allow the following operations to be carried out for free: (i) to apply any unitary operation that commutes with the total Hamiltonian, (ii) to bring in any ancillary system in its Gibbs state at inverse temperature β, and (iii) to discard any system. The most general transformation a system S can undergo under this set of rules is a *thermal operation* defined using an additional system B with a Hamiltonian H_B and a unitary U_{SB} with $[U_{SB}, H_S + H_B] = 0$, as

$$\Phi_S(\cdot) = \mathrm{tr}_B\Big[U_{SB}\left((\cdot)\otimes\gamma_B\right)U_{SB}^{\dagger}\Big]\,, \tag{5}$$

where $\gamma_B = e^{-\beta H_B}/\,\mathrm{tr}(e^{-\beta H_B})$ is the Gibbs state of the bath system B. Our second main theorem echoes Theorem I using the framework of thermal operations, yet our proof is restricted to processes that are time-covariant, i.e., that commute with the time evolution. (The assumption of time covariance allows us to sidestep issues of coherence between energy levels [33–37].) Specifically, we construct a universal protocol for implementing any time-covariant i.i.d. channel using thermal operations and a battery, at a work cost per copy that is asymptotically equal to the thermodynamic capacity.

Theorem IV. *Let X be a system with Hamiltonian H_X, and let $\beta \geqslant 0$. Let $\mathcal{E}_X$ be any completely positive, trace-preserving map that satisfies, for all t,*

$$\mathcal{E}_X(e^{-iH_Xt}\,(\cdot)\,e^{iH_Xt}) = e^{-iH_Xt}\,\mathcal{E}_X(\cdot)\,e^{iH_Xt}\,. \tag{6}$$

Then, for $\epsilon \in (0,1]$ there exists a thermal operation acting on $X^{\otimes n}$ and a work storage system, whose induced mapping Φ_{X^n} on $X^{\otimes n}$ is as $\frac{1}{2}\|\Phi_{X^n} - \mathcal{E}_X^{\otimes n}\|_{\diamond} \leqslant \epsilon$, with the amount of work $W_{\mathrm{th}}^{\epsilon}(\mathcal{E}^{\otimes n})$ invested satisfying

$$\lim_{n\to\infty}\frac{1}{n}W_{\mathrm{th}}^{\epsilon}(\mathcal{E}^{\otimes n}) = T(\mathcal{E})\,. \tag{7}$$

This work cost is optimal.

This directly implies that Corollary III also holds in the context of thermal operations for time-covariant processes: such processes may be reversibly inter-converted in the i.i.d. regime using thermal operations and a battery. The proof of Theorem IV, presented in Appendix G, proceeds by exhibiting an explicit energy-conserving unitary acting on the system, work storage system and a large bath, that achieves the desired transformation. A main step is to adapt recently developed ideas from quantum information theory, including the convex-split lemma and position-based decoding [25–29, 38].

Extensions.—We show that these ideas also provide a one-shot conditional erasure protocol that is valid for systems described by a non-trivial Hamiltonian and states that are time-covariant (Appendix G.2), thus generalizing the protocol of Ref. [9]. The work cost is given in terms of the hypothesis testing entropy D_H^{ϵ} defined in (Appendix A). This information measure extends the standard quantum relative entropy to the one-shot regime [39–44], and is closely-related to other one-shot information measures [45–47]. Our result implies that it is possible to implement any time-covariant process for a fixed time-covariant input state in the single-shot regime, using thermal operations and a battery, at a cost given by the coherent relative entropy.

Proposition V. *Let S, M be quantum systems with Hamiltonians H_S, H_M, and let R be a reference system. For any $|\rho\rangle_{SMR}$ such that $[\rho_{SM}, H_S + H_M] = 0$ and $\epsilon \in (0,1]$, there exists a thermal operation acting on S, M and a battery system, that maps*

$$\rho_{SMR} \mapsto \gamma_S \otimes \rho_{MR}\,, \tag{8}$$

where $\gamma_S = e^{-\beta H_S}/\,\mathrm{tr}(e^{-\beta H_S})$, and that consumes an amount of work that is approximately $D_{\mathrm{H}}^{1-\epsilon}(\rho_{SM}\,\|\,\gamma_S \otimes \rho_M)$.

Finally, we show if the input is a fixed i.i.d. state, it is possible to implement any arbitrary, not necessarily time-covariant i.i.d. channel using thermal operations, a battery, and a sub-linear amount of coherence, at the same asymptotic work cost per copy as it would take to implement it with Gibbs-preserving maps (Appendix H). We thus conclude that although Gibbs-preserving maps are more powerful in general than thermal operations [48], they become asymptotically equivalent in the macroscopic limit in terms of implementing i.i.d. processes on given i.i.d. input states.

Discussion.—Whereas the interconversion of quantum states in the various resource theories of quantum thermodynamics is relatively well understood, the interconversion of channels has received less attention. We show that in the i.i.d. regime, the situation for channels is similar to that of quantum states. Asymptotically, there exists a unique monotone, the thermodynamic capacity, which characterizes the "value" of a channel. In this sense our result is the thermodynamic analogue of the reverse Shannon theorem for quantum channels. Note, however, that in the information-theoretic setting the strong assumption of free entanglement is needed [13, 32]. In contrast, we consider our thermodynamic setting to be very natural. Importantly, the thermodynamic reversibility described here refers to conversion between the channels themselves, and not between the input and the output of a specific channel. In general, it is not possible to recover the input state of a channel from its output with a universal protocol, since the thermodynamic capacity is computed with a maximization over the inputs of the channel.

Our statements and proofs are also fully noncommutative in the sense that they cannot be simplified to a problem about classical probability distributions in a fixed basis—a feature that is still rather uncommon in quantum thermodynamics. Moreover, standard proof techniques developed for quantum channel simulations do not readily apply to our problems at hand. For example, a natural attempt at proving Theorem I would be to mimic the logic of Refs. [32, 49], and to exploit properties of the relevant entropy measures. In fact, this proof strategy works for systems described by a trivial Hamiltonian $H = 0$. There, the expression (3) for i.i.d. channels can be reduced to a single conditional max-entropy quantity using the post-selection technique [15]. Then, exploiting a quasi-convexity-like property of the conditional max-entropy allows to prove our result for the special case of trivial Hamiltonians (Appendix D). Attempting to generalize this proof approach to non-trivial Hamiltonians fails because the coherent relative entropy does not display the required quasi-convexity property for non-trivial Hamiltonians.

The thermodynamic capacity exhibits some interesting elementary properties (Appendix B). It can be computed explicitly, and analytically for some simple examples, as it can be formulated as a convex optimization problem [50]. The thermodynamic capacity is also additive [10], and does not need to be regularized as for other channel capacity measures.

Outlook.—Whether or not it is possible to implement any i.i.d. channel that is not time-covariant using thermal operations is still an open question. We expect that such a protocol might in general need a very large amount of coherence, much like the requirement of large embezzling states for the reverse Shannon theorem [13, 32]. Indeed, if the input is a superposition of two different i.i.d. states of different energy, the environment must be able to coherently compensate for any energy difference caused by the process without disturbing the process. However, as we have seen for fixed i.i.d. input states any i.i.d. channel can be implemented optimally using thermal operations. In that sense, studying how to implement any i.i.d. channel using thermal operations will shine further light on the differences and similarities of Gibbs-preserving and thermal maps [48].

Given the relevance of one-shot entropy measures for a wide range of physical and information-theoretic situations, we expect our results to find applications beyond thermodynamic interconversion of processes. For instance, we note that a quantity closely related to the coherent relative entropy has found applications in studying dissipative dynamics of many-body systems [51]. Also, in contrast to standard smooth entropy measures for quantum states [47], our channel smoothing in terms of the diamond norm leaves one of the marginals invariant when applied to quantum states (cf. the very recent related works [52, 53]). This might offer some insights on the quantum joint typicality conjecture in quantum communication theory [21, 54, 55], on which recent process has been made [24].

Finally, that there exists optimal universal thermodynamic implementations of channels indicates that low-dissipation components for future quantum devices can in principle be developed, that function accurately for all inputs, and still dissipate no more than required by the worst case input.

Acknowledgments.—We thank Álvaro Alhambra, David Ding, Patrick Hayden, Rahul Jain and David Jennings for discussions. PhF acknowledges support from the Swiss National Science Foundation (SNSF) through the Early PostDoc.Mobility Fellowship No. P2EZP2_165239 hosted by the Institute for Quantum Information and Matter (IQIM) at Caltech, from the IQIM which is a National Science Foundation (NSF) Physics Frontiers Center (NSF Grant PHY-1733907), and from the Department of Energy Award DE-SC0018407. FB is supported by the NSF.

APPENDICES

The appendices are structured as follows. Appendix A introduces the necessary preliminaries and fixes some notation. In Appendix B we introduce the thermodynamic capacity and calculate the thermodynamic capacity of some simple example channels. Appendix C then reviews some additional tools that we need based on Schur-Weyl duality, such as universal entropy and energy expectation value estimation, as well as the postselection technique. In Appendix D, we consider the case of systems described by a trivial Hamiltonian, and we prove our main result in this situation by using a reasoning similar to refs. [32, 49]. We also show that this reasoning fails in the general case, because the coherent relative entropy does not display even approximately the required quasi-convexity property. Appendix E is devoted to the proof of our main result in the context of Gibbs-sub-preserving maps, specifically proving Proposition II and Theorem I of the main text. As an interlude, Appendix F provides a related result, namely a new proof of the asymptotic equipartition property of the coherent relative entropy that is significantly simpler and provides better insight than the proof in Ref. [6]. Appendix G concerns our second main result. There, we construct a collection of protocols for conditional Landauer erasure using side information (see Fig. 3), which we then use to prove Theorem IV as well as the related result Proposition V of the main text. In Appendix H we provide an implementation of any i.i.d. process for a fixed i.i.d. input state, using thermal operations, a battery and a small amount of coherence. Finally, some technical lemmas are collected in Appendix I.

Appendix A: Preliminaries and notation

Preliminaries. Each quantum system considered lives in a finite-dimensional Hilbert space. A quantum state is positive semidefinite operator ρ satisfying $\mathrm{tr}(\rho) = 1$. A subnormalized quantum state is a positive semidefinite operator ρ satisfying $\mathrm{tr}(\rho) \leqslant 1$. To each system S is associated a standard basis, usually denoted by $\{|k\rangle_S\}$. For any two systems A, A', we denote by $A \simeq A'$ the fact that they are isometric; in that case we consider a representation in which the isometry maps the standard basis onto the standard basis, i.e., $\mathrm{id}_{A\to A'}(|k\rangle\langle k|_A) = |k\rangle\langle k|_{A'}$ for all k, where $\mathrm{id}_{A\to A'}$ denotes the identity process. For any two systems $A \simeq A'$, we define the nonnormalized maximally entangled reference ket $|\Phi\rangle_{A:A'} = \sum_k |k\rangle_A \otimes |k\rangle_{A'}$. Matrix inequalities are with respect to the positive semidefinite cone: $A \leqslant B$ signifies that $B - A$ is positive semidefinite. A completely positive map $\mathcal{E}_{X\to X'}$ is a linear mapping that maps Hermitian operators on X to Hermitian operators on X' and that satisfies $\mathcal{E}_{X\to X'}(\Phi_{X:R_X}) \geqslant 0$. The adjoint $\mathcal{E}^\dagger_{X\leftarrow X'}$ of a completely positive map $\mathcal{E}_{X\to X'}$ is the unique completely positive map $X' \to X$ that satisfies $\mathrm{tr}(\mathcal{E}(Y)\, Z) = \mathrm{tr}(Y \mathcal{E}^\dagger(Z))$ for all operators Y, Z. A completely positive map $\mathcal{E}_{X\to X'}$ is trace-preserving if $\mathcal{E}^\dagger(\mathbb{1}_{X'}) = \mathbb{1}_X$ and trace-nonincreasing if $\mathcal{E}^\dagger(\mathbb{1}_{X'}) \leqslant \mathbb{1}_X$.

Proximity of quantum states can be measured by the fidelity $F(\rho, \sigma) = \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1$, where the one-norm of an operator is defined as $\left\|A\right\|_1 = \mathrm{tr}\,\sqrt{A^\dagger A}$. An associated metric can be defined as $P(\rho, \sigma) = \sqrt{1 - \bar{F}^2(\rho, \sigma)}$, called the purified distance [47, 56, 57], or root infidelity, and is closely related to the Bures distance and the quantum angle [58]. (Our definition of the purified distance differs slightly from that of refs. [47, 56, 57], in that we base it on the fidelity function and not on a generalization of the fidelity to subnormalized states. The two definitions of the purified distance yet coincide as long as at least one of the two states is normalized. Still, the properties presented here hold in full generality for subnormalized states as well, and the present measure proves slightly more convenient to deal with in our work.) The proximity of two subnormalized quantum states ρ, σ may also be measured in the trace distance $D(\rho, \sigma) := \frac{1}{2}\left\|\rho - \sigma\right\|_1$. We note that the one-norm of a Hermitian operator A can be expressed as

$$\|A\|_1 = \max_{\|Z\|_\infty \leqslant 1} \operatorname{tr}(ZA) = \min_{\substack{\Delta_\pm \geqslant 0 \\ A = \Delta_+ - \Delta_-}} \operatorname{tr}(\Delta_+) + \operatorname{tr}(\Delta_-) \ , \tag{A.1}$$

where the first optimization ranges over Hermitian Z operators and where the second over positive semidefinite operators $\Delta_\pm$. For any two subnormalized quantum states ρ, σ, the purified distance and the trace distance are related via

$$D(\rho, \sigma) \leqslant P(\rho, \sigma) \leqslant \sqrt{2D(\rho, \sigma)} \ . \tag{A.2}$$

Similarly, we may define a distance measure for channels: For two completely positive, trace-nonincreasing maps $\mathcal{T}_{X \to X'}$ and $\mathcal{T}'_{X \to X'}$, the diamond norm distance is defined as

$$\frac{1}{2} \|\mathcal{T}_{X \to X'} - \mathcal{T}'_{X \to X'}\|_\diamond = \max_{\sigma_{XR}} D\big(\mathcal{T}_{X \to X'}(\sigma_{XR}), \mathcal{T}'_{X \to X'}(\sigma_{XR})\big) \ , \tag{A.3}$$

where the optimization ranges over all bipartite quantum states over X and a reference system $R \simeq X$. The optimization may be restricted to pure states without loss of generality.

Entropy measures. The von Neumann entropy of a quantum state ρ is

$$S(\rho) = -\operatorname{tr}(\rho \ln \rho) \ . \tag{A.4}$$

In this work, all entropies are defined in units of nats, using the natural logarithm $\ln(\cdot)$, instead of units of (qu)bits. A number of nats is equal to $\ln(2)$ times the corresponding number of qubits. The conditional von Neumann entropy of a bipartite state ρ_{AB} is given by

$$S(A \,|\, B)_\rho = S(AB)_\rho - S(B)_\rho = S(\rho_{AB}) - S(\rho_B) \ . \tag{A.5}$$

The quantum relative entropy is defined as

$$D(\rho \,\|\, \tau) = \operatorname{tr}\big[\rho\big(\ln \rho - \ln \tau\big)\big] \ , \tag{A.6}$$

where ρ is a quantum state and where τ is any positive semidefinite operator whose support contains the support of ρ. Finally, one of our proofs rely on the hypothesis testing entropy [39–43] in its form as presented in [44]. It given by the following equivalent optimizations, which are semidefinite programs [59], in terms of the primal variable $Q \geqslant 0$ and the dual variables $\mu, X \geqslant 0$:

$$\begin{aligned} \exp\{-D_{\mathrm{H}}^{\eta}(\rho \,\|\, \Gamma)\} \quad &= \quad \begin{array}{rl} \text{minimize}: & \eta^{-1} \operatorname{tr}(Q\Gamma) \\ \text{subject to}: & Q \leqslant \mathbb{1} \\ & \operatorname{tr}(Q\rho) \geqslant \eta \ ; \end{array} & \text{(A.7)} \\ &= \quad \begin{array}{rl} \text{maximize}: & \mu - \eta^{-1} \operatorname{tr}(X) \\ \text{subject to}: & \mu\rho \leqslant \Gamma + X \ . \end{array} & \text{(A.8)} \end{aligned}$$

Thermodynamic framework. We consider the framework of Ref. [6]. To each system S considered is associated a positive semidefinite operator $\Gamma_S \geqslant 0$. A trace-nonincreasing, completely positive map $\Phi_{A \to B}$ is allowed for free if it satisfies $\Phi_{A \to B}(\Gamma_A) \leqslant \Gamma_B$.

The resources required to enable forbidden operations are counted using an explicit system that provides these resources, such as an *information battery.* An information battery is a large register that is in a state of the special form $\tau_m = P_m/\operatorname{tr}(P_m)$ where P_m is a projector of rank e^m. I.e., τ_m has uniform eigenvalues over a given rank e^m. The basic resource here is purity: We choose measure it in pure nats, equal to $\ln(2)$ times a number of pure qubits. Here, the purity of τ_m is simply $\ln(d) - m$, where d is the dimension of the information battery. A useful characterization of which processes can be implemented using an information battery is given by the following proposition.

Proposition 1 ([6, Proposition II]). *Let $\Gamma_X, \Gamma_{X'} \geqslant 0$. Let $\mathcal{T}_{X\to X'}$ be a completely positive, trace-nonincreasing map. Then $\mathcal{T}_{X\to X'}$ may be implemented using free operations and an information battery while expending a work cost equivalent to λ pure nats if, and only if,*

$$\mathcal{T}_{X\to X'}(\Gamma_X) \leqslant e^{\lambda}\,\Gamma_{X'} \ . \tag{A.9}$$

The resources can be counted in terms of thermodynamic work in units of energy if we are given a heat bath at inverse temperature T. Recall that a pure qubit can be converted to $kT\ln(2)$ work using a Szilárd engine, where k is Boltzmann's constant [60]. By counting purity in nats instead of qubits, we get rid of the $\ln(2)$ factor: A number λ of pure nats can be converted into $\lambda\, kT$ thermodynamic work using a Szilárd-type engine.

In these appendices we count work exclusively in equivalent of pure nats, for simplicity, as opposed to units of energy as in the main text. The two are directly related by a factor $\beta^{-1} = kT$. Furthermore, this eliminates the factor β from otherwise essentially information-theoretic expressions, and our theorems thus directly apply to cases where $\Gamma_X, \Gamma_{X'}$ are any abstract positive semidefinite operators which are not necessarily defined via a Hamiltonian.

In Ref. [6], the resource cost λ of implementing a process $\mathcal{E}_{X\to X'}$ (any completely positive, trace-preserving map) up to an accuracy $\epsilon \geqslant 0$ in terms of proximity of the process matrix given a fixed input state σ_X, counted in pure nats, was shown to be given by the coherent relative entropy

$$\lambda = -\hat{D}^{\epsilon}_{X\to X'}(\mathcal{E}_{X\to X'}(\sigma_{XR_X}) \,\|\, \Gamma_X, \Gamma_{X'}) = \ln \min_{\substack{\mathcal{T}(\Gamma_X)\leqslant \alpha\Gamma_{X'}\\ P(\mathcal{T}(\sigma_{XR_X}),\mathcal{E}(\sigma_{XR_X}))\leqslant\epsilon}} \alpha \ , \tag{A.10}$$

where σ_{XR_X} is the purification of σ_X on a system $R_X \simeq X$ given by $|\sigma\rangle_{XR} = \sigma_X^{1/2}\,|\Phi\rangle_{X:R_X}$, and where the optimization ranges over completely positive, trace-nonincreasing maps $\mathcal{T}_{X\to X'}$. The coherent relative entropy enjoys a collection of properties in relation to the conditional min- and max-entropy, and to the min- and max-relative entropy. It satisfies the following asymptotic equipartition property, for any completely positive, trace-preserving $\mathcal{E}_{X\to X'}$, for any quantum state σ_X and for any $0 < \epsilon < 1$:

$$\lim_{n\to\infty} \frac{1}{n}\hat{D}^{\epsilon}_{X^n\to X'^n}\big(\mathcal{E}^{\otimes n}_{X\to X'}(\sigma^{\otimes n}_{XR}) \,\big\|\, \Gamma^{\otimes n}_X, \Gamma^{\otimes n}_{X'}\big) = D(\sigma_X \,\|\, \Gamma_X) - D(\mathcal{E}(\sigma_X) \,\|\, \Gamma_{X'}) \ . \tag{A.11}$$

A further interesting property is that for any $\mathcal{E}_{X\to X'}$, for any σ_X and for $\epsilon = 0$, we have [6, Proposition 31]

$$\hat{D}_{X\to X'}(\mathcal{E}_{X\to X'}(\sigma_{XR_X}) \,\|\, \Gamma_X, \Gamma_{X'}) \leqslant D(\sigma_X \,\|\, \Gamma_X) - D(\mathcal{E}(\sigma_X) \,\|\, \Gamma_{X'}) \ . \tag{A.12}$$

(A corresponding inequality for $\epsilon > 0$ can be derived using the continuity of the quantum relative entropy in its first argument.) Here, we are interested in the corresponding quantity that is independent of the input state. Namely, as we can see from Proposition 1, the resource cost of universally implementing a process $\mathcal{E}_{X\to X'}$ for any input state, and measured in pure nats, is given by the quantity

$$W^{\epsilon}_{X\to X'}(\mathcal{E}_{X\to X'} \,\|\, \Gamma_X, \Gamma_{X'}) = \ln \min_{\substack{\mathcal{T}(\Gamma_X)\leqslant \alpha\,\Gamma_{X'} \\ \frac{1}{2}\|\mathcal{T}-\mathcal{E}\|_\diamond \leqslant \epsilon}} \alpha \ , \tag{A.13}$$

where again $\mathcal{T}_{X\to X'}$ ranges over all trace-nonincreasing, completely positive maps. The condition on the diamond norm ensures that $\mathcal{T}$ is indeed a universal implementation of the process, i.e., that $\mathcal{T}$ accurately reproduces the output of the ideal process for any possible input including inputs that are correlated with a reference system.

Appendix B: The thermodynamic capacity

The *thermodynamic capacity* of a completely positive, trace-preserving map $\mathcal{E}_{X\to X'}$ relative to operators $\Gamma_X, \Gamma_{X'} \geqslant 0$ is defined by

$$T(\mathcal{E}) = \sup_{\sigma_X} \Big\{ D(\mathcal{E}_{X\to X'}(\sigma_X) \,\|\, \Gamma_{X'}) - D(\sigma_X \,\|\, \Gamma_X) \Big\} \ , \tag{B.1}$$

which we express here in number of pure nats instead of usual units of energy, as discussed above, effectively setting $\beta = 1$. We may rewrite the argument in the optimization (B.1) as

$$\begin{aligned} &D(\mathcal{E}_{X\to X'}(\sigma_X) \,\|\, \Gamma_{X'}) - D(\sigma_X \,\|\, \Gamma_X) \\ &\qquad = -S(\mathcal{E}_{X\to X'}(\sigma_X)) + S(\sigma_X) - \mathrm{tr}(\mathcal{E}_{X\to X'}(\sigma_X)\ \ln \Gamma_{X'}) + \mathrm{tr}(\sigma_X\ \ln \Gamma_X) \\ &\qquad = S(E\,|\,X')_\rho - \mathrm{tr}(\mathcal{E}_{X\to X'}(\sigma_X)\ \ln \Gamma_{X'}) + \mathrm{tr}(\sigma_X\ \ln \Gamma_X)\ , \end{aligned} \tag{B.2}$$

where we have defined the state $\rho_{EX'} = V_{X\to X'}\sigma_X V^\dagger$ using a Stinespring dilation isometry $V_{X\to X'E}$ of $\mathcal{E}_{X\to X'}$ into an environment system E, satisfying $\mathcal{E}_{X\to X'}(\cdot) = \mathrm{tr}_E\big(V\,(\cdot)\,V^\dagger\big)$. Note that the conditional entropy is concave in the quantum state. This is because $S(E\,|\,X')_\rho = -D(\rho_{EX'} \,\|\, \mathbb{1}_E \otimes \rho_{X'})$, and the quantum relative entropy is jointly convex. Hence, the optimization (B.1) is a convex optimization that can be carried out efficiently for small system sizes [50]. Indeed, we have successfully computed the thermodynamic capacity of simple example quantum channels acting on few qubits with Python code, using the QuTip framework [61, 62] and the CVXOPT optimization software [63].

The thermodynamic capacity is also additive [10]. Intuitively, an implementation of many copies of $\mathcal{E}$ is also an implementation of half as many copies of $\mathcal{E}^{\otimes 2}$, so we expect the thermodynamic capacity to display this property.

Proposition 2 (Additivity of the thermodynamic capacity [10])**.** *For any $\Gamma_X, \Gamma_{X'}, \Gamma_Z, \Gamma_{Z'} \geqslant 0$, and for any quantum channels $\mathcal{E}_{X\to X'}$, $\mathcal{F}_{Z\to Z'}$, we have $T(\mathcal{E}\otimes\mathcal{F}) = T(\mathcal{E}) + T(\mathcal{F})$.*

Proof of Proposition 2. Let σ_X, τ_Z be states achieving the thermodynamic capacity of $T(\mathcal{E})$ and $T(\mathcal{F})$, respectively. Then $\sigma_X \otimes \tau_Z$ is a candidate for $T(\mathcal{E} \otimes \mathcal{F})$, yielding

$$\begin{aligned} T(\mathcal{E} \otimes \mathcal{F}) &\geqslant D(\mathcal{E}(\sigma) \otimes \mathcal{F}(\tau) \,\|\, \Gamma_{X'} \otimes \Gamma_{Z'}) - D(\sigma \otimes \tau \,\|\, \Gamma_X \otimes \Gamma_Z) \\ &= D(\mathcal{E}(\sigma) \,\|\, \Gamma_{X'}) - D(\sigma \,\|\, \Gamma_X) + D(\mathcal{F}(\tau) \,\|\, \Gamma_{Z'}) - D(\tau \,\|\, \Gamma_Z) \\ &= T(\mathcal{E}) + T(\mathcal{F}) \ . \end{aligned} \tag{B.3}$$

Now let ζ_{XZ} achieve the optimum for $T(\mathcal{E} \otimes \mathcal{F})$. Let $V_{X \to E_1 X'}$, $W_{Z \to E_2 Z'}$ be Stinespring isometries of $\mathcal{E}$ and $\mathcal{F}$ respectively, such that $\mathcal{E}(\cdot) = \mathrm{tr}_{E_1}\big(V\,(\cdot)\,V^\dagger\big)$ and $\mathcal{F}(\cdot) = \mathrm{tr}_{E_2}\big(W\,(\cdot)\,W^\dagger\big)$. Let $\rho_{E_1 E_2 X' Z'} = (V \otimes W)\,\zeta\,(V \otimes W)^\dagger$. Then

$$\begin{aligned} T(\mathcal{E} \otimes \mathcal{F}) &= D((\mathcal{E} \otimes \mathcal{F})(\zeta) \,\|\, \Gamma_{X'} \otimes \Gamma_{Z'}) - D(\zeta_{XZ} \,\|\, \Gamma_X \otimes \Gamma_Z) \\ &= S(E_1 E_2 \,|\, X'Z')_\rho - \mathrm{tr}[\rho_{X'Z'} \ln(\Gamma_{X'} \otimes \Gamma_{Z'})] + \mathrm{tr}[\zeta_{XZ} \ln(\Gamma_X \otimes \Gamma_Z)] \ , \\ &= S(E_1 E_2 \,|\, X'Z')_\rho - \mathrm{tr}[\rho_{X'} \ln(\Gamma_{X'})] - \mathrm{tr}[\rho_{Z'} \ln(\Gamma_{Z'})] \\ &\quad + \mathrm{tr}[\zeta_X \ln(\Gamma_X)] + \mathrm{tr}[\zeta_Z \ln(\Gamma_Z)] \ , \end{aligned} \tag{B.4}$$

since $\ln(A \otimes B) = \ln(A) \otimes \mathbb{1} + \mathbb{1} \otimes \ln(B)$. Invoking the chain rule of the von Neumann entropy, and then strong subadditivity of the entropy, we see that $S(E_1 E_2 \,|\, X'Z')_\rho = S(E_1 \,|\, X'Z')_\rho + S(E_2 \,|\, E_1 X'Z')_\rho \leqslant S(E_1 \,|\, X')_\rho + S(E_2 \,|\, Z')_\rho$. Hence

$$\begin{aligned} \text{(B.4)} &\leqslant S(E_1 \,|\, X')_\rho - \mathrm{tr}[\rho_{X'} \ln(\Gamma_{X'})] + \mathrm{tr}[\zeta_X \ln(\Gamma_X)] \\ &\qquad + S(E_2 \,|\, Z')_\rho - \mathrm{tr}[\rho_{Z'} \ln(\Gamma_{Z'})] + \mathrm{tr}[\zeta_Z \ln(\Gamma_Z)] \\ &\leqslant T(\mathcal{E}) + T(\mathcal{F}) \ , \end{aligned} \tag{B.5}$$

where the last inequality holds because the reduced states ζ_X, ζ_Z are optimization candidates for $T(\mathcal{E})$ and $T(\mathcal{F})$, respectively. ■

We now calculate the thermodynamic capacity explicitly for some simple examples.

Example: The identity channel. Let $X' \simeq X$. Let $\Gamma_X, \Gamma_{X'} \geqslant 0$, where $\Gamma_{X'}$ might differ from Γ_X. The thermodynamic capacity of the identity channel $\mathrm{id}_{X \to X'}$ becomes

$$T(\mathrm{id}_{X \to X'}) = \max_\sigma \{\mathrm{tr}(\sigma \ln \Gamma_X) - \mathrm{tr}(\sigma \ln \Gamma_{X'})\} = \max_\sigma \mathrm{tr}[\sigma(\ln \Gamma_X - \ln \Gamma_{X'})] \ . \tag{B.6}$$

Hence, the thermodynamic capacity picks out the maximal eigenvalue of $\ln \Gamma_X - \ln \Gamma_{X'}$. If we think of $-\ln \Gamma$ as a Hamiltonian, the thermodynamic capacity picks out the worst possible change in energy between input and output. (The same holds for any unitary channel, as long as we ensure that the input and output Γ operators are suitably rotated with respect to each other.)

Example: The constant channel. A constant channel has the form $\mathcal{E}_{X \to X'}(\cdot) = \mathrm{tr}(\cdot)\,\rho_{X'}$, destroying the input and creating a fixed output $\rho_{X'}$ regardless of the input. Given $\Gamma_X, \Gamma_{X'} \geqslant 0$, for any σ_X we have

$$D(\mathcal{E}(\sigma_X) \,\|\, \Gamma_{X'}) - D(\sigma_X \,\|\, \Gamma_X) = D(\rho_{X'} \,\|\, \Gamma_{X'}) - D(\sigma_X \,\|\, \Gamma_X) \ . \tag{B.7}$$

The maximum of this expression is attained when $D(\sigma_X \,\|\, \Gamma_X)$ is minimal, which occurs when $\sigma_X = \Gamma_X / \mathrm{tr}(\Gamma_X)$. Intuitively, this is because the state that minimizes the free energy is the thermal state; alternatively, write $D(\sigma_X \,\|\, \Gamma_X) = D(\sigma_X \,\|\, \Gamma_X / \operatorname{tr}(\Gamma_X)) - \ln \operatorname{tr}(\Gamma_X)$, and we know that the quantum relative entropy, for two normalized states, is always positive and is equal to zero if and only if the two arguments are equal. Hence,

$$\begin{aligned} T(\mathcal{E}) &= D(\rho_{X'} \,\|\, \Gamma_{X'}) - \ln \operatorname{tr}(\Gamma_X) \\ &= -S(\rho_{X'}) - \operatorname{tr}(\rho_{X'} \ln \Gamma_{X'}) - \ln \operatorname{tr}(\Gamma_X) \ . \end{aligned} \tag{B.8}$$

Special case: Trivial Hamiltonians. Another special case that is worth mentioning is the case where $\Gamma_X = \mathbb{1}_X$, $\Gamma_{X'} = \mathbb{1}_{X'}$, which corresponds to the situation where the Hamiltonians of X and X' are trivial. For any quantum channel $\mathcal{E}_{X\to X'}$, let $V_{X\to X'E}$ be a Stinespring dilation isometry satisfying $\mathcal{E}_{X\to X'}(\cdot) = \operatorname{tr}_E\big(V\,(\cdot)\,V^\dagger\big)$. Then we have

$$T(\mathcal{E}) = \sup_\sigma \{S(\sigma_X) - S(\mathcal{E}(\sigma_X))\} = \sup_\sigma S(E \,|\, X')_{V\sigma V^\dagger} \ . \tag{B.9}$$

That is, the thermodynamic capacity characterizes by how much the channel is capable of reducing the entropy of its input, or equivalently, how much entropy the channel is capable of dumping into the environment when conditioned on the output.

Appendix C: Prelude on Schur-Weyl duality and the postselection technique

In order to prove Lemma 13, we need some additional tools from quantum information theory which are based on Schur-Weyl duality, and heavily inspired by the techniques of refs. [12–14, 21, 24]. First, we note that the measurement of the isotypic component of n-tensor powers is a good estimate of the entropy of a n-tensor state. Then, we note that a similar measurement is a good estimate for the total energy of the n-tensor state. These two observations combined together will allow us to construct our universal relative and conditional typical selection operator. Finally, we need the postselection technique to bound the diamond norm of channels over n systems [15, 32].

C.1. Schur-Weyl duality, definitions and notation

Consider a Hilbert space $\mathscr{H}_A$ and let $n \in \mathbb{N}$. The group $\mathrm{GL}(d_A) \times \mathrm{S}_n$ acts naturally on $\mathscr{H}_A^{\otimes n}$, where $X \in \mathrm{GL}(d_A)$ acts as $X^{\otimes n}$ and where the permutation group permutes the tensor factors. We mostly follow the notation of refs. [12, 14]. Schur-Weyl tells us that the full Hilbert space decomposes as

$$\mathscr{H}_A \simeq \bigoplus_\lambda \mathcal{V}_\lambda = \bigoplus_\lambda \mathcal{Q}_\lambda \otimes \mathcal{P}_\lambda \ , \tag{C.1}$$

where $\lambda \in \mathrm{Young}(n, d)$ are Young diagrams with n boxes and (at most) d rows, and where $\mathcal{Q}_\lambda$, $\mathcal{P}_\lambda$ are irreducible representations of $\mathrm{GL}(d_A)$ and S_n, respectively. The number of Young diagrams in the decomposition above is at most $\mathrm{poly}(n)$, if d_A is kept constant. We denote by $\Pi^\lambda_{A^n}$ the projector in $\mathscr{H}_A^{\otimes n}$ onto the term labeled by λ in the decomposition above. We denote by $q_\lambda(X)$ a representing matrix of $X \in \mathrm{GL}(d_A)$ in the irreducible representation labeled by λ; the operator $q_\lambda(X)$ lives in $\mathcal{Q}_\lambda$. We furthermore introduce the notation, for any $Y \in \mathcal{Q}_\lambda \otimes \mathcal{P}_\lambda$,

$$[\, Y \,]_\lambda \;= \mathbb{1}_{(\mathcal{Q}_\lambda \otimes \mathcal{P}_\lambda) \to A^n}\; Y \; \mathbb{1}^\dagger_{(\mathcal{Q}_\lambda \otimes \mathcal{P}_\lambda) \leftarrow A^n} \tag{C.2}$$

the canonical embedding of an operator Y on $\mathcal{Q}_\lambda \otimes \mathcal{P}_\lambda$ into the space $\mathscr{H}_A^{\otimes n}$, i.e., mapping Y onto the corresponding block in (C.1). In particular,

$$\Pi^\lambda_{A^n}\, [\, Y \,]_\lambda \, \Pi^\lambda_{A^n} = \,[\, Y \,]_\lambda \;\; . \tag{C.3}$$

An important result is the following: Any operator X_{A^n} acting on the n copies which commutes with all the permutations admits a decomposition of the form

$$X_{A^n} = \sum [\, X_\lambda \otimes \mathbb{1}_{\mathcal{P}_\lambda} \,]_\lambda \;\; , \tag{C.4}$$

for some set of operators $X_\lambda \in \mathcal{Q}_\lambda$. In particular, $[X_{A^n}, \Pi^\lambda_{A^n}] = 0$. We can make this more precise for i.i.d. states. For any $X \in \mathrm{GL}(d_A)$, we have that

$$[\Pi^\lambda_{A^n}, X^{\otimes n}] = 0 \;, \tag{C.5}$$

$$X^{\otimes n} = \sum_\lambda [\, q_\lambda(X) \otimes \mathbb{1}_{\mathcal{P}_\lambda} \,]_\lambda \;\; . \tag{C.6}$$

For a given $\lambda \in \mathrm{Young}(n, d)$, it is often useful to consider the corresponding normalized probability distribution $\lambda/n = (\lambda_i/n)_i$. The entropy of this distribution is given by

$$\bar{S}(\lambda) := S(\lambda/n) = -\sum_i \frac{\lambda_i}{n} \ln \frac{\lambda_i}{n} \;, \tag{C.7}$$

where λ_i is the number of boxes in the i-th row of the diagram.

A useful expression for $\Pi^\lambda_{A^n}$ may be obtained following [14, Section V]. We have

$$\begin{aligned} \Pi^\lambda_{A^n} &= \frac{\dim(\mathcal{Q}_\lambda)}{s_\lambda(\mathrm{diag}(\lambda/n))} \int dU_A \, \Pi^\lambda_{A^n} \Big(U_A \; \mathrm{diag}(\lambda/n)_A \, U_A^\dagger \Big)^{\otimes n} \, \Pi^\lambda_{A^n} \\ &\leqslant \mathrm{poly}(n)\, e^{n\bar{S}(\lambda)} \int dU_A \, \Big(U_A \; \mathrm{diag}(\lambda/n)_A \, U_A^\dagger \Big)^{\otimes n} \;, \end{aligned} \tag{C.8}$$

recalling that $\Pi^\lambda_{A^n}$ commutes with any i.i.d. state, with $s_\lambda(X) = \mathrm{tr}(q_\lambda(X))$ and using bounds on $\dim(\mathcal{Q}_\lambda)$ and $s_\lambda(\mathrm{diag}(\lambda/n))$ derived in Ref. [14]. Here dU_A denotes the Haar measure over all unitaries acting on $\mathscr{H}_A$, normalized such that $\int dU_A = 1$.

If we have n copies of a bipartite system $\mathscr{H}_A \otimes \mathscr{H}_B$, then we may Schur-Weyl decompose $\mathscr{H}_A^{\otimes n}$, $\mathscr{H}_B^{\otimes n}$ and $(\mathscr{H}_A \otimes \mathscr{H}_B)^{\otimes n}$ under the respective actions of $\mathrm{GL}(d_A) \times \mathrm{S}_n$, $\mathrm{GL}(d_B) \times \mathrm{S}_n$ and $\mathrm{GL}(d_A d_B) \times \mathrm{S}_n$. A useful property we will need here is that the projectors onto the respective Schur-Weyl blocks commute between these decompositions.

Lemma 3. *Consider two spaces* $\mathscr{H}_A, \mathscr{H}_B$. *Let* $\Pi^\lambda_{(AB)^n}$ *and* $\Pi^{\lambda'}_{A^n}$ *be the projectors onto Schur-Weyl blocks of* $\mathscr{H}_{AB}^{\otimes n}$ *and* $\mathscr{H}_A^{\otimes n}$, *respectively, with* $\lambda \in \mathrm{Young}(d_A d_B, n)$ *and* $\lambda' \in \mathrm{Young}(d_A, n)$.

Then

$$[\Pi^{\lambda}_{(AB)^n}, \Pi^{\lambda'}_{A^n} \otimes \mathbb{1}_{B^n}] = 0 \ . \tag{C.9}$$

Proof of Lemma 3. $\Pi^{\lambda'}_{A^n} \otimes \mathbb{1}_{B^n}$ is invariant under the action of S_n permuting the copies of $A \otimes B$, and so it admits a decomposition of the form (C.4) and commutes with $\Pi^{\lambda}_{(AB)^n}$. ■

Here is also another lemma about "how much overlap" Schur-Weyl blocks have on a bipartite system versus on one of the two systems. This lemma forms the basis of our universal typical subspace, inspired by earlier approaches [12–14, 21, 24].

Lemma 4. *Consider n copies of a bipartite system $\mathscr{H}_A \otimes \mathscr{H}_B$. Then for any $\lambda \in \mathrm{Young}(d_A d_B, n)$ and $\lambda' \in \mathrm{Young}(d_B, n)$, we have*

$$\Pi^{\lambda'}_{B^n}\ \mathrm{tr}_{A^n}\Big(\Pi^{\lambda}_{(AB)^n}\Big)\,\Pi^{\lambda'}_{B^n} \leqslant \mathrm{poly}(n)\, e^{n(\bar{S}(\lambda)-\bar{S}(\lambda'))}\,\Pi^{\lambda'}_{B^n}\ , \tag{C.10}$$

noting that $[\mathbb{1}_{A^n} \otimes \Pi^{\lambda'}_{B^n}, \Pi^{\lambda}_{(AB)^n}] = 0$.

Proof of Lemma 4. Using the expression (C.8) for $\Pi^{\lambda}_{(AB)^n}$, we have

$$\mathrm{tr}_{A^n}\Big(\Pi^{\lambda}_{(AB)^n}\Big) \leqslant \mathrm{poly}(n)\, e^{n\bar{S}(\lambda)} \int dU_{AB}\ \mathrm{tr}_{A^n}\Big[\Big(U_{AB}\ \mathrm{diag}(\lambda/n)_{AB}\, U^{\dagger}_{AB}\Big)^{\otimes n}\Big]\ . \tag{C.11}$$

Observe that for any state τ_B, we have

$$\begin{aligned} \big\|\Pi^{\lambda'}_{B^n}\, \tau_B^{\otimes n}\, \Pi^{\lambda'}_{B^n}\big\|_\infty = \big\|\ [\ q_{\lambda'}(\tau_B) \otimes \mathbb{1}_{\mathcal{P}_{\lambda'}}\ ]_{\lambda'}\ \big\|_\infty = \|q_{\lambda'}(\tau_B)\|_\infty \leqslant \mathrm{tr}(q_{\lambda'}(\tau_B)) \\ \leqslant \mathrm{poly}(n)\, e^{-n\bar{S}(\lambda')}\ , \end{aligned} \tag{C.12}$$

as derived e.g. in Ref. [14], and thus for any state τ_B,

$$\Pi^{\lambda'}_{B^n}\, \tau_B^{\otimes n}\, \Pi^{\lambda'}_{B^n} \leqslant \mathrm{poly}(n)\, e^{-n\bar{S}(\lambda')}\, \Pi^{\lambda'}_{B^n}\ . \tag{C.13}$$

Hence,

$$\begin{aligned} &\Pi^{\lambda'}_{B^n}\ \mathrm{tr}_{A^n}\Big(\Pi^{\lambda}_{(AB)^n}\Big)\,\Pi^{\lambda'}_{B^n} \\ &\qquad \leqslant \mathrm{poly}(n)\, e^{n\bar{S}(\lambda)} \int dU_{AB}\ \Pi^{\lambda'}_{B^n}\ \Big(\mathrm{tr}_A\Big(U_{AB}\ \mathrm{diag}(\lambda/n)_{AB}\, U^{\dagger}_{AB}\Big)\Big)^{\otimes n}\, \Pi^{\lambda'}_{B^n} \\ &\qquad \leqslant \mathrm{poly}(n)\, e^{n\bar{S}(\lambda)} \int dU_{AB}\ \ \mathrm{poly}(n)\, e^{-n\bar{S}(\lambda')}\, \Pi^{\lambda'}_{B^n} \\ &\qquad = \mathrm{poly}(n)\, e^{n(\bar{S}(\lambda)-\bar{S}(\lambda'))}\, \Pi^{\lambda'}_{B^n}\ , \end{aligned} \tag{C.14}$$

as required. ■

C.2. Estimating the entropy with a universal observable

Interestingly, measuring the Young diagram λ (i.e., performing the projective measurement with operators $\{\Pi^{\lambda}_{A^n}\}_{\lambda}$) yields a pretty good estimation of the spectrum of a state ρ_A when given $\rho_A^{\otimes n}$ [14]. An estimate for the entropy of ρ is thus obtained by calculating the entropy $S(\lambda/n)$ corresponding to probability distribution λ/n. More precisely, for any state ρ_A and for any $\delta > 0$, we have that (see, e.g., [12, Eq. (6.23)]):

$$\mathrm{tr}\left\{\left(\sum_{\lambda:\ \bar{S}(\lambda)\in[S(\rho)\pm\delta]}\Pi^{\lambda}_{A^n}\right)\rho_A^{\otimes n}\right\} \geqslant 1-\mathrm{poly}(n)\exp\{-n\eta\}\ , \tag{C.15}$$

where $\eta > 0$ is given as a function of δ and d_A but is independent of n, λ, ρ_A. (More precisely, we need to invert the Fannes-Audenaert continuity bound to deduce from $|S(\lambda/n) - S(\rho)| > \delta$ that $\|\lambda/n - \mathrm{eig}(\rho)\|_1 > \delta'$ for some $\delta' > 0$, then let $\eta := \delta'^2/2 > 0$.)

C.3. Estimating the energy expectation value

Proposition 5. *Consider any observable H_A on $\mathscr{H}_A$. We write $\Gamma_A = e^{-H_A}$. Then the set of projectors $\left\{R^k_{A^n}\right\}$ onto the eigenspaces of $\Gamma_A^{\otimes n}$ forms a POVM satisfying the following properties:*

(i) There are at most poly(n) *POVM elements, with the label k running over a set $k \in \mathcal{K}_n(H_A) \subset \mathbb{R}$;*

(ii) We have $[R^k_{A^n}, \Gamma_A^{\otimes n}] = 0$ and $e^{-nk}\, R^k_{A^n} = R^k_{A^n}\,\Gamma_A^{\otimes n}$;

(iii) For any $\delta > 0$ and for any state ρ_A,

$$\mathrm{tr}\left\{R^{\approx_\delta \mathrm{tr}(\rho H)}_{A^n}\,\rho_A^{\otimes n}\right\} \geqslant 1 - 2\,e^{-n\eta}\ , \tag{C.16}$$

where we have defined for any $h \in \mathbb{R}$

$$R^{\approx_\delta h}_{A^n} = \sum_{k\in\mathcal{K}_n(H_A)\,:\,|k-h|\leqslant\delta} R^k_{A^n}\ , \tag{C.17}$$

and where $\eta = \delta^2/(2\|H_A\|_\infty^2)$;

(iv) For any $h \in \mathbb{R}$, we have $e^{-n(k+\delta)}R^{\approx_\delta h}_{A^n} \leqslant R^{\approx_\delta h}_{A^n}\,\Gamma_A^{\otimes n} \leqslant e^{-n(k-\delta)}R^{\approx_\delta h}_{A^n}$.

Proof of Proposition 5. The fact that there are only poly(n) elements follows because there are only so many types. Property (ii) holds by definition. Property (iv) holds because $e^{-n(k\pm\delta)}$ is the minimum/maximum eigenvalue of $\Gamma_A^{\otimes n}$ in the subspace spanned by $R^{\approx_\delta h}_{A^n}$. Finally, we need to show Property (iii): This follows from a large deviation analysis. More precisely, let Z_j for $j = 1, \ldots, n$ be random variables where Z_j represents the measurement outcome of H_A on the j-th system of the i.i.d. state $\rho_A^{\otimes n}$. By Hoeffding's inequality, we have that

$$\Pr\left[\left|(1/n)\sum Z_j - \mathrm{tr}(\rho_A H_A)\right| > \delta\right] \leqslant 2\exp\left(-\frac{2n\delta^2}{\Delta H_A^2}\right) \leqslant 2\exp\left(-\frac{n\delta^2}{2\,\|H_A\|_\infty^2}\right)\ , \tag{C.18}$$

where ΔH_A is the difference between the maximum and minimum eigenvalue of H_A, and $\Delta H_A \leqslant 2\|H_A\|_\infty$. Thus the event consisting of the outcomes k satisfying $|k - \mathrm{tr}(\rho_A H_A)| \leqslant \delta$ happens with probability at least $1 - 2e^{-n\eta}$, proving (C.16). ■

C.4. Postselection technique for quantum channels

The postselection technique is useful for bounding the diamond norm of a candidate smoothed channel to a target ideal i.i.d. channel.

Theorem 6 (Postselection technique [15, 32])**.** *Let X, X' be quantum systems, and let $\mathcal{E}_{X\to X'}$ be a completely positive, trace-preserving map. Let $\mathcal{T}_{X^n\to X'^n}$ be a completely positive, trace-nonincreasing map. Let $\bar{R} \simeq X$ and let*

$$\tau_{X^n} := \mathrm{tr}_{\bar{R}^n}\left\{ \int d\phi_{X\bar{R}}\, |\phi\rangle\langle\phi|_{X\bar{R}}^{\otimes n} \right\} = \int d\sigma_X\, \sigma_X^{\otimes n} \ , \tag{C.19}$$

where $d\phi_{X\bar{R}}$ denotes the Haar-induced measure on the pure states on $X \otimes \bar{R}$, and $d\sigma_X$ its induced measure on X after partial trace. Let $|\tau\rangle_{X^n R}$ be a purification of τ_{X^n}. Then

$$\frac{1}{2}\|\mathcal{T} - \mathcal{E}^{\otimes n}\|_\diamond \leqslant \mathrm{poly}(n)\, D\big(\mathcal{T}(\tau_{X^n R}), \mathcal{E}^{\otimes n}(\tau_{X^n R})\big) \ . \tag{C.20}$$

Furthermore, for all n there exists a set $\{|\phi_i\rangle_{X\bar{R}}\}$ of at most $\mathrm{poly}(n)$ states, and a probability distribution $\{p_i\}$, providing a purification of τ_{X^n} as

$$|\tau\rangle_{X^n \bar{R}^n R'} = \sum_i \sqrt{p_i}\, |\phi_i\rangle_{X\bar{R}}^{\otimes n} \otimes |i\rangle_{R'} \ , \tag{C.21}$$

with a register R' of size $\mathrm{poly}(n)$.

Now we present a convenient proposition which allows to prove that a given channel is close to an i.i.d. channel, if it behaves as expected (including with the correct global phase) on all i.i.d. states with exponentially good accuracy.

Proposition 7. *For three systems X, X', E, let $V_{X\to X'E}$ be an isometry, and let $W_{X^n\to X'^n E^n}$ be an isometry which commutes with the permutations of the n systems. Assume that there exists $\eta > 0$ independent of n such that for all states $|\sigma\rangle_{XR_X}$ with a reference system $R_X \simeq X$, we have*

$$\mathrm{Re}\left\{ \langle\sigma|_{XR_X}^{\otimes n} (V_{X\leftarrow X'E}^\dagger)^{\otimes n}\, W_{X^n\to X'^n E^n}\, |\sigma\rangle_{XR_X}^{\otimes n} \right\} \geqslant 1 - \mathrm{poly}(n)\exp(-n\eta) \ . \tag{C.22}$$

Let $\mathcal{E}_{X\to X'}(\cdot) = \mathrm{tr}_E\big(V_{X\to X'E}\, (\cdot)\, V^\dagger\big)$ and $\mathcal{T}_{X^n\to X'^n}(\cdot) = \mathrm{tr}_{E^n}\big(W_{X^n\to X'^n E^n}\, (\cdot)\, W^\dagger\big)$. Then

$$\frac{1}{2}\big\|\mathcal{T}_{X^n\to X'^n} - \mathcal{E}_{X\to X'}^{\otimes n}\big\|_\diamond \leqslant \mathrm{poly}(n)\exp(-n\eta/2) \ . \tag{C.23}$$

Proof of Proposition 7. We use the postselection technique above to bound the diamond norm distance between $\mathcal{T}_{X^n\to X'^n}$ and $\mathcal{E}^{\otimes n}_{X\to X'}$. Let $|\tau\rangle_{X^n\bar{R}^n R'}$ be the purification of the de Finetti state given by (C.21). Calculate

$$\begin{aligned}
&\mathrm{Re}\big\{\langle\tau|_{X^n\bar{R}^n R'}(V^{\otimes n}_{X\to EX'})^\dagger W_{X^n\to E^nX'^n}|\tau\rangle_{X^n\bar{R}^n R'}\big\}\\
&\qquad=\sum p_i\ \mathrm{Re}\Big\{\langle\phi_i|^{\otimes n}_{X\bar{R}}\,(V^{\otimes n}_{X\to EX'})^\dagger\,W_{X^n\to E^nX'^n}|\phi_i\rangle^{\otimes n}_{X\bar{R}}\Big\}\\
&\qquad\geqslant 1-\mathrm{poly}(n)\exp(-n\eta)\ ,
\end{aligned}\tag{C.24}$$

which implies, recalling that $F(|\psi\rangle,|\phi\rangle)=|\langle\psi|\phi\rangle|\geqslant\mathrm{Re}\{\langle\psi|\phi\rangle\}$ and that $(1-x)^2\geqslant 1-2x$,

$$F^2\big(V^{\otimes n}_{X\to EX'}\,|\tau\rangle_{X^n\bar{R}^n R'}\,,\,W_{X^n\to E^nX'^n}|\tau\rangle_{X^n\bar{R}^n R'}\big)\geqslant 1-\mathrm{poly}(n)\exp(-n\eta)\ ,\tag{C.25}$$

and hence

$$P\big(V^{\otimes n}_{X\to EX'}\,|\tau\rangle_{X^n\bar{R}^n R'}\,,\,W_{X^n\to E^nX'^n}|\tau\rangle_{X^n\bar{R}^n R'}\big)\leqslant \mathrm{poly}(n)\exp(-n\eta/2)\ .\tag{C.26}$$

Recalling the relations between the trace distance and the purified distance, and noting that these distance measures cannot increase under the partial trace, we obtain

$$\begin{aligned}
&D\big(\mathcal{T}(\tau_{X^n\bar{R}^n R'}),\mathcal{E}^{\otimes n}(\tau_{X^n\bar{R}^n R'})\big)\leqslant P\big(\mathcal{T}(\tau_{X^n\bar{R}^n R'})\ ,\ \mathcal{E}^{\otimes n}(\tau_{X^n\bar{R}^n R'})\big)\\
&\qquad\leqslant P\big(W_{X^n\to E^nX'^n}|\tau\rangle_{X^n\bar{R}^n R'}\ ,\ V^{\otimes n}_{X\to EX'}\,|\tau\rangle_{X^n\bar{R}^n R'}\big)\leqslant\mathrm{poly}(n)\exp(-n\eta/2)\ .
\end{aligned}\tag{C.27}$$

The post-selection technique then asserts that

$$\frac{1}{2}\|\mathcal{T}-\mathcal{E}^{\otimes n}\|_\diamond\leqslant\mathrm{poly}(n)\ \exp(-n\eta/2)\ ,\tag{C.28}$$

as claimed. ∎

Interestingly, in the above proof it does not matter that the number of terms in the de Finetti state decomposition (C.21) is at most $\mathrm{poly}(n)$.

Appendix D: Approach using properties of the relevant one-shot entropy measures

The difficult direction of our main theorem in the special case where $\Gamma_X=\mathbb{1}_X$, $\Gamma_{X'}=\mathbb{1}_{X'}$ can be shown using an argument similar to that of refs. [32, 49], by exploiting a quasi-convexity property of the max-entropy. By using the post-selection technique, and recalling that the fixed-input state case is given by the coherent relative entropy, we find

$$W^{\epsilon}_{X^n\to X'^n}\big(\mathcal{E}^{\otimes n}_{X\to X'}\,\big\|\,\mathbb{1}_{X^n},\mathbb{1}_{X'^n}\big)\leqslant -\hat{D}^{\epsilon/\mathrm{poly}(n)}_{X^n\to X'^n}\big(\mathcal{E}^{\otimes n}_{X\to X'}(\tau_{X^nR^n_X})\,\big\|\,\mathbb{1}_{X^n},\mathbb{1}_{X'^n}\big)\ .\tag{D.1}$$

(Indeed, the optimal implementation given by the coherent relative entropy is a valid implementation for the problem defining the left hand side.) In the case of trivial Hamiltonians, the coherent relative entropy reduces to the smooth max-entropy [6]; more precisely

$$\hat{D}^{\epsilon}_{X\to X'}\big(\rho_{X'R_X}\,\big\|\,\mathbb{1}_X,\mathbb{1}_{X'}\big) \geqslant -H^{c\epsilon^{\alpha}}_{\max}(E\,|\,X')_{\rho} + g(\epsilon)\ , \tag{D.2}$$

where $|\rho\rangle_{X'R_XE}$ is a pure state, where $c>0$, $0<\alpha<1$, $g(\epsilon)$ are universal and do not depend on the state or the dimensions of the systems, and where the smooth max-entropy is defined as

$$H^{\epsilon}_{\max}(E\,|\,X')_{\rho} = \min_{P(\hat{\rho},\rho)\leqslant\epsilon} H^{\epsilon}_{\max}(E\,|\,X')_{\hat{\rho}}\ ; \tag{D.3}$$

$$H_{\max}(E\,|\,X')_{\hat{\rho}} = \max_{\omega_{X'}} \ln \big\|\hat{\rho}^{1/2}_{EX'}\omega^{1/2}_{X'}\big\|_1^2\ , \tag{D.4}$$

and thus

$$\text{(D.1)} \leqslant H^{\epsilon^{\alpha}/\mathrm{poly}(n)}_{\max}(E^n\,|\,X'^n)_{\rho} + g(\epsilon)\ , \tag{D.5}$$

where $\rho_{X'^nE^n} = V^{\otimes n}_{X\to X'E}\tau_{X^n}(V^{\dagger})^{\otimes n} = \int d\sigma\,(V\sigma V^{\dagger})^{\otimes n}$, where $V_{X\to X'E}$ is a Stinespring dilation isometry of $\mathcal{E}_{X\to X'}$ as $\mathcal{E}_{X\to X'}(\cdot) = \mathrm{tr}_E\big(V_{X\to X'E}\,(\cdot)\,V^{\dagger}\big)$. At this point we invoke two facts. First, note that the de Finetti state can be written as a mixture of only poly(n) i.i.d. states, instead of a continuous average (Theorem 6): There exists a set $\{\sigma_i\}$ of at most poly(n) states and a distribution $\{p_i\}$ such that $\tau_{X^n} = \sum_i p_i\sigma_i^{\otimes n}$. Second, we invoke the fact the conditional max-entropy is quasi-convex up to a penalty term, namely, that the conditional max-entropy of $\sum_i p_i\rho_i$ is less than or equal to the maximum over the set of max-entropies corresponding to each ρ_i, plus a term proportional to the logarithm of the number of terms in the sum. We prove this as Proposition 8 below. Hence, with $\rho_i = V\,\sigma_i\,V^{\dagger}$, we have

$$\text{(D.5)} \leqslant \max_i \hat{H}^{\epsilon^{\alpha}/\mathrm{poly}(n)}_{\max}(E^n\,|\,X'^n)_{\rho_i^{\otimes n}} + \ln(\mathrm{poly}(n)) + g(\epsilon)\ . \tag{D.6}$$

Now we are in business because the max-entropy is evaluated on an i.i.d. state, and we know it asymptotically goes to the von Neumann entropy in this regime. Also, $\lim_{n\to\infty}(1/n)[\ln(\mathrm{poly}(n)) + g(\epsilon)] = 0$ and hence

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}W^{\epsilon}_{X^n\to X'^n}\big(\mathcal{E}^{\otimes n}_{X\to X'}\,\big\|\,\mathbb{1}_{X^n},\mathbb{1}_{X'^n}\big) &\leqslant \max_i\ S(E\,|\,X')_{\rho_i}\\ &= \max_i\{S(\sigma_i) - S(\mathcal{E}(\sigma_i))\}\\ &\leqslant \max_\sigma\{S(\sigma) - S(\mathcal{E}(\sigma))\}\\ &= T(\mathcal{E})\ ,\end{aligned} \tag{D.7}$$

noting that $S(E\,|\,X') = S(EX') - S(X') = S(X) - S(X')$ and recalling that we are considering the special case where $\Gamma_X = \mathbb{1}_X, \Gamma_{X'} = \mathbb{1}_{X'}$. This completes the proof of the difficult direction of our main theorem in the special case of a trivial system Hamiltonian.

Above, we used the fact that the max-entropy is quasi-convex up to a penalty term, which we prove here:

Proposition 8. *Let $\{\rho^i_{AB}\}$ be a set of M quantum states, and let $\{p_i\}$ be a probability* *distribution. Let* $\rho_{AB} = \sum_{i=1}^{M} p_i \rho^i_{AB}$*. Then*

$$H^{\epsilon}_{\max}(A\,|\,B)_{\rho} \leqslant \max_i H^{\epsilon}_{\max}(A\,|\,B)_{\rho^i} + \ln(M^2)\ . \tag{D.8}$$

Proof of Proposition 8. Let $\{\hat\rho^i_{AB}\}$ be states that achieve the smooth max-entropy for each i, that is, $H^{\epsilon}_{\max}(A\,|\,B)_{\rho^i} = H_{\max}(A\,|\,B)_{\hat\rho^i}$. Let $\hat\rho_{AB} = \sum p_i \hat\rho^i_{AB}$. Observe that $P(\hat\rho_{AB}, \rho_{AB}) \leqslant \epsilon$, which follows from e.g. [32, Lemma A.3]. Now let $\hat\omega_B$ achieve the maximum for the candidate smoothed max-entropy $H_{\max}(A\,|\,B)_{\hat\rho} = 2\ln F(\hat\rho_{AB}, \mathbb{1}_A\otimes\hat\omega_B)$. By invoking [64, Lemma 4.9], which states that $F(\sum A_i, B) \leqslant \sum F(A_i, B)$, we may assert that

$$F(\hat\rho_{AB}, \mathbb{1}_A\otimes\hat\omega_B) \leqslant \sum \sqrt{p_i}\, F(\hat\rho^i_{AB}, \mathbb{1}_A\otimes\hat\omega_B) \leqslant M\ \max_i F(\hat\rho^i_{AB}, \mathbb{1}_A\otimes\hat\omega_B)\ . \tag{D.9}$$

Then,

$$\begin{aligned} H^{\epsilon}_{\max}(A\,|\,B)_{\rho} &\leqslant H_{\max}(A\,|\,B)_{\hat\rho} = 2\ln F(\hat\rho_{AB}, \mathbb{1}_A\otimes\hat\omega_B) \\ &\leqslant 2\ln(M) + \max_i\ 2\ln F(\hat\rho^i_{AB}, \mathbb{1}_A\otimes\hat\omega_B) \\ &\leqslant 2\ln(M) + \max_i H_{\max}(A\,|\,B)_{\hat\rho^i} \\ &= 2\ln(M) + \max_i H_{\max}(A\,|\,B)_{\rho^i}\ , \end{aligned} \tag{D.10}$$

which completes the proof. ∎

Naturally, one might ask whether it is possible to extend this proof to the case of nontrivial Γ operators. Interestingly, this is not possible, at least not in a naive way. The problem is that we need a quasi-convexity property of the form

$$\begin{aligned} &-\hat D^{\epsilon}_{X\to X'}\big(\mathcal{E}_{X\to X'}(\sigma_{XR_X})\,\big\|\,\Gamma_X, \Gamma_{X'}\big) \\ &\qquad\overset{?}{\leqslant} \max_i\Big(-\hat D^{\epsilon}_{X\to X'}\big(\mathcal{E}_{X\to X'}(\sigma^i_{XR_X})\,\big\|\,\Gamma_X, \Gamma_{X'}\big)\Big) + (\text{penalty})\ ,\ (?) \end{aligned} \tag{D.11}$$

where $\sigma_X = \sum p_i\sigma^i_X$ and $|\sigma\rangle_{XR} = \sigma_X^{1/2}\,|\Phi\rangle_{X:R_X}$, $|\sigma^i\rangle_{XR} = (\sigma^i_X)^{1/2}\,|\Phi\rangle_{X:R_X}$, and where the (penalty) term scales in a favorable way in n, say of order $\ln(\mathrm{poly}(M))$ where M is the number of terms in the convex decomposition as for the max-entropy. In fact, Eq. (D.11) is false, as can be shown using an explicit counter-example on a two-level system. As the counter-example is based on physical reasons, the coherent relative entropy is not even approximately quasi-convex. We note that this counter-example doesn't exclude a quasi-convexity property that might have a penalty term that depends on properties of the Γ operators, yet such a term would likely scale unfavorably with n.

Consider a two-level system with a Hamiltonian H with two energy levels $|0\rangle, |1\rangle$ at corresponding energies $E_0 = 0$ and $E_1 > 0$. The corresponding Γ operator is $\Gamma = g_0|0\rangle\langle 0| + g_1|1\rangle\langle 1|$ with $g_0 = 1$, $g_1 = e^{-\beta E_1}$. Consider the process consisting in erasing the input and creating the output state $|+\rangle$, where we define $|\pm\rangle = [|0\rangle \pm |1\rangle]/\sqrt{2}$. That is, we consider the process $\mathcal{E}(\cdot) = \mathrm{tr}(\cdot)\,|+\rangle\langle +|$. Suppose the input state is maximally mixed, $\sigma = \mathbb{1}/2$, such that $\rho_{X'R_X} = |+\rangle\langle +|_{X'}\otimes\mathbb{1}_{R_X}/2$. If $E_0 = 0$ and $E_1 \to \infty$, then this process requires a lot of work; intuitively, with probability $1/2$ we start in the ground state $|0\rangle$ and need to prepare the output state $|+\rangle$ which has high energy. For $\epsilon = 0$, we can see this because the input state is full rank, hence $\mathcal{T} = \mathcal{E}$; then $\mathcal{E}(\Gamma) = \operatorname{tr}(\Gamma)|+\rangle\langle+|$ and the smallest α such that $\mathcal{E}(\Gamma) \leqslant \alpha\Gamma$ is given by $\alpha/\operatorname{tr}(\Gamma) = \|\Gamma^{-1/2}|+\rangle\langle+|\Gamma^{-1/2}\|_\infty = \langle+|\Gamma^{-1}|+\rangle = (g_0^{-1} + g_1^{-1})/2 = (1 + e^{\beta E_1})/2 \geqslant e^{\beta E_1}/2$. Noting that $\operatorname{tr}(\Gamma) \geqslant 1$, we have $\alpha \geqslant e^{\beta E_1}/2$, and hence the energy cost of the transformation $\mathbb{1}/2 \to |+\rangle$ is

$$\text{energy cost} = -\beta^{-1}\hat{D}_{X\to X'}(\rho_{X'R_X} \,\|\, \Gamma, \Gamma) = \beta^{-1}\ln\alpha \geqslant E_1 - \beta^{-1}\ln(2)\ . \tag{D.12}$$

Clearly, this work cost can become arbitrarily large if $E_1 \to \infty$. On the other hand, we can perform the transformation $|+\rangle \to |+\rangle$ obviously at no work cost; similarly, $|-\rangle \to |+\rangle$ can be carried out by letting the system time-evolve under its own Hamiltonian for exactly the time interval required to pick up a relative phase (-1) between the $|0\rangle$ and $|1\rangle$ states. This also costs no work because it is a unitary operation that commutes with the Hamiltonian. We thus have our counter-example to the quasi-convexity of the coherent relative entropy. The transformation $\mathbb{1}/2 \to |+\rangle$ is very hard, but the individual transformations $|\pm\rangle \to |+\rangle$ are trivial, noting that $\mathbb{1}/2 = (1/2)|+\rangle\langle+| + (1/2)|-\rangle\langle-|$.

We can also make the above claim robust against an accuracy tolerance $\epsilon \geqslant 0$. First, notice that the condition $P(\mathcal{T}_{X\to X'}(\sigma_{XR_X}), \rho_{X'R_X}) \leqslant \epsilon$ implies that $\frac{1}{2}\|\mathcal{T}_{X\to X'}(\sigma_{XR_X}) - \rho_{X'R_X}\|_1 \leqslant \epsilon$, which in turn implies that $(1/4)\left\|\mathcal{T}_{X\to X'}(\Phi_{X:R_X}) - |+\rangle\langle+|_{X'} \otimes \mathbb{1}_{R_X}\right\|_1 \leqslant \epsilon$, and hence that $\mathcal{T}_{X\to X'}(\cdot) = \operatorname{tr}(\cdot)\,|+\rangle\langle+|_{X'} + \Delta(\cdot)$ for some Hermiticity-preserving map $\Delta(\cdot)$ satisfying $\frac{1}{2}\|\Delta(\Phi_{XR_X})\|_1 \leqslant 2\epsilon$. Then $\mathcal{T}_{X\to X'}(\Gamma) \leqslant \alpha\Gamma$ implies that $\alpha\Gamma \geqslant \operatorname{tr}(\Gamma)|+\rangle\langle+| + \Delta(\Gamma) \geqslant |+\rangle\langle+| - \Delta_-$ for $\Delta_- \geqslant 0$ defined as the negative part of $\Delta(\Gamma)$, satisfying $\operatorname{tr}(\Delta_-) \leqslant 4\epsilon$, and since $\operatorname{tr}(\Gamma) \geqslant 1$. Hence, $\alpha^{-1}\,|+\rangle\langle+| \leqslant \Gamma + \Delta_-/\alpha$. Hence for any $0 < \eta \leqslant 1$ to be fixed later, $\mu = \alpha^{-1}$ is feasible for the dual problem defining the hypothesis testing entropy $D_{\mathrm{H}}^{\eta}(|+\rangle\langle+| \,\|\, \Gamma)$, and $e^{-D_{\mathrm{H}}^{\eta}(|+\rangle\langle+|\,\|\,\Gamma)} \geqslant \alpha^{-1} - \operatorname{tr}(\Delta_-/\alpha)/\eta \geqslant \alpha^{-1} - \alpha^{-1}(4\epsilon/\eta)$. Thus $\ln(\alpha) \geqslant D_{\mathrm{H}}^{\eta}(|+\rangle\langle+| \,\|\, \Gamma) + \ln(1 - 4\epsilon/\eta)$. Choosing $\eta = 8\epsilon$ yields $\ln(1 - 4\epsilon/\eta) = -\ln(2)$. On the other hand, by definition of the hypothesis testing entropy we have that $e^{-D_{\mathrm{H}}^{\eta}(|+\rangle\langle+|\,\|\,\Gamma)} \leqslant \operatorname{tr}(Q\Gamma)/\eta$ for any $0 \leqslant Q \leqslant \mathbb{1}$ satisfying $\operatorname{tr}(Q|+\rangle\langle+|) \geqslant \eta$; with $Q = 2\eta|1\rangle\langle1|$ we obtain $e^{-D_{\mathrm{H}}^{\eta}(|+\rangle\langle+|\,\|\,\Gamma)} \leqslant 2g_1 = 2e^{-\beta E_1}$. Then, $\ln(\alpha) \geqslant -\ln(2) + \beta E_1 - \ln(2) = -2\ln(2) + \beta E_1$. For the optimal α of the coherent relative entropy, we finally see that the transformation $\mathbb{1}/2 \to |+\rangle$ may require arbitrarily much energy if $E_1 \to \infty$, even for a small $\epsilon > 0$:

$$\text{energy cost} = -\beta^{-1}\hat{D}^{\epsilon}_{X\to X'}(\rho_{X'R_X} \,\|\, \Gamma, \Gamma) = \beta^{-1}\ln(\alpha) \geqslant E_1 - 2\beta^{-1}\ln(2)\ . \tag{D.13}$$

Appendix E: Proof of our main result

This appendix is dedicated to the proof of Theorem I and Proposition II in the main text. The proof of Theorem I is split in two parts, corresponding to each inequality direction. The first direction is straightforward (Appendix E.1), while the other direction is less so (Appendix E.2 and Appendix E.3).

E.1. Easy direction

The easy direction of our main theorem essentially states that a universal implementation must be itself a valid candidate special-purpose implementation for any input state, and must thus require at least that much work. We provide a brief formal proof of this relatively obvious statement.

Lemma 9. *Let $\Gamma_X, \Gamma_{X'} \geqslant 0$, and let $\mathcal{E}_{X\to X'}$ be any completely positive, trace-preserving map. Then, for any quantum state σ_X, and any $\epsilon \geqslant 0$, we have*

$$W^{\epsilon}_{X\to X'}(\mathcal{E}_{X\to X'} \,\|\, \Gamma_X, \Gamma_{X'}) \geqslant -\hat{D}^{\sqrt{2\epsilon}}_{X\to X'}(\mathcal{E}_{X\to X'}(\sigma_{XR_X}) \,\|\, \Gamma_X, \Gamma_{X'}) \ , \tag{E.1}$$

where σ_{XR_X} is a standard purification of σ_X. Moreover,

$$\lim_{n\to\infty} \frac{1}{n} W^{\epsilon}_{X^n\to X'^n}(\mathcal{E}^{\otimes n}_{X^n\to X'^n} \,\|\, \Gamma_X^{\otimes n}, \Gamma_{X'}^{\otimes n}) \geqslant \max_{\sigma_X} \left[D(\mathcal{E}(\sigma_X) \,\|\, \Gamma_{X'}) - D(\sigma_X \,\|\, \Gamma_X) \right] \ . \tag{E.2}$$

Proof of Lemma 9. Let $\mathcal{T}$ satisfy $\frac{1}{2}\|\mathcal{E} - \mathcal{T}\|_\diamond \leqslant \epsilon$. Let σ_X be any quantum state, and let $|\sigma\rangle_{XR_X} = \sigma_X^{1/2} |\Phi\rangle_{X:R_X}$. Then by definition of the diamond norm it must hold that $D\big(\mathcal{E}(\sigma_{XR_X}), \mathcal{T}(\sigma_{XR_X})\big) \leqslant \epsilon$, which implies that $P\big(\mathcal{E}(\sigma_{XR_X}), \mathcal{T}(\sigma_{XR_X})\big) \leqslant \sqrt{2\epsilon}$. Then $\mathcal{T}$ is a valid optimization candidate for the definition of the coherent relative entropy, and (E.1) follows. Furthermore, for any σ_X, we have

$$\begin{aligned}
\lim_{n\to\infty} &\frac{1}{n} W^{\epsilon}_{X^n\to X'^n}(\mathcal{E}^{\otimes n}_{X^n\to X'^n} \,\|\, \Gamma_X^{\otimes n}, \Gamma_{X'}^{\otimes n}) \\
&\geqslant -\lim_{n\to\infty} \frac{1}{n} \hat{D}^{\sqrt{2\epsilon}}_{X^n\to X'^n}(\mathcal{E}^{\otimes n}_{X\to X'}(\sigma^{\otimes n}_{XR_X}) \,\|\, \Gamma_X^{\otimes n}, \Gamma_{X'}^{\otimes n}) \\
&= D(\mathcal{E}(\sigma_X) \,\|\, \Gamma_{X'}) - D(\sigma_X \,\|\, \Gamma_X) \ ,
\end{aligned} \tag{E.3}$$

where the last inequality follows from the asymptotic equipartition property (A.11) of the coherent relative entropy. The relation above holds for all σ_X, thus establishing (E.2). ∎

E.2. Difficult direction: Trivial Hamiltonians first

As a warm-up, we first consider the case of a trivial input and output Hamiltonian, i.e., $\Gamma_X = \mathbb{1}_X$ and $\Gamma_{X'} = \mathbb{1}_{X'}$. The present simpler situation illustrates the use of a universal conditional typical projector to smooth a n-tensor process, allowing to establish the i.i.d. behavior of $W^{\epsilon}_{X\to X'}(\mathcal{E}_{X\to X'} \,\|\, \mathbb{1}_X, \mathbb{1}_{X'})$. The proof presented in this section is redundant, in that a different proof applying to the same situation was already given in Appendix D. We provide the present alternative proof in order to get some additional intuition about a possible form of a (near) optimal process, and because it forms the basis for our generalization to the case $\Gamma \neq \mathbb{1}$.

Here we show explicitly that there is a universal implementation $\mathcal{T}_{X^n\to X'^n}$ of $\mathcal{E}^{\otimes n}_{X\to X'}$ such that, for any input state σ_X, we have $\mathcal{T}_{X^n\to X'^n}(\sigma^{\otimes n}_{XR_X}) \approx \mathcal{E}^{\otimes n}(\sigma^{\otimes n}_{XR_X})$ with exponentially small error in n; furthermore the work cost rate of $\mathcal{T}_{X^n\to X'^n}$ is arbitrarily close to $\max_\sigma[S(\sigma_X) - S(\mathcal{E}(\sigma_X))]$.

First we construct a universal typical subspace for the quantum conditional entropy, based on ideas from Schur-Weyl duality. This will be useful for our main proof. The construction presented here is similar to, and heavily inspired by, techniques put forward in earlier work [12–14, 21, 24].

Proposition 10. *Consider $\mathscr{H}_A, \mathscr{H}_B$ and let $S \in \mathbb{R}$. For any $\delta > 0$, there exists a projector $P^{S,\delta}_{A^nB^n}$ acting on $(\mathscr{H}_A \otimes \mathscr{H}_B)^{\otimes n}$, and there exists $\eta > 0$, such that:*

(i) $P^{S,\delta}_{A^nB^n}$ *is permutation-invariant;*

(ii) For any quantum state ρ_{AB} with $S(A|B)_\rho \leqslant S$,

$$\mathrm{tr}\big(P^{S,\delta}_{A^nB^n}\,\rho^{\otimes n}_{AB}\big) \geqslant 1 - \mathrm{poly}(n)\exp(-n\eta)\ ; \tag{E.4}$$

(iii) $\mathrm{tr}_{A^n}\big(P^{S,\delta}_{A^nB^n}\big) \leqslant \mathrm{poly}(n)\, e^{n(S+2\delta)}\, \mathbb{1}_{B^n}$.

To understand why this projector is "conditional," and for a simple illustration of its use, consider the smooth Rényi-zero conditional max-entropy, also known as the smooth alternative max-entropy. It is defined for a bipartite state ρ_{AB} as

$$\hat{H}^{\epsilon}_{\max}(A\,|\,B)_\rho = \min_{\hat{\rho}\approx_\epsilon \rho} \ln \big\|\mathrm{tr}_A(\Pi^{\hat{\rho}_{AB}}_{AB})\big\|_\infty\ , \tag{E.5}$$

where $\Pi^{\hat{\rho}_{AB}}_{AB}$ is the projector onto the support of $\hat{\rho}_{AB}$, and where the optimization ranges over subnormalized states $\hat{\rho}_{AB}$ which are ϵ-close to ρ_{AB} in purified distance. We may understand the i.i.d. behavior of this quantity as follows. Let $\delta > 0$, and for any n let $P^{S,\delta}_{(AB)^n}$ be the projector constructed by Proposition 10 for the value $S = S(A\,|\,B)_\rho$. Define $\hat{\rho}_{A^nB^n} = P^{S,\delta}\,\rho^{\otimes n}_{AB}\,P^{S,\delta}$. Then $\hat{\rho}_{A^nB^n} \approx_\epsilon \rho^{\otimes n}_{AB}$ for n large enough, thanks to Property (ii) and the gentle measurement lemma (e.g., Lemma 27). On the other hand, using Property (iii),

$$\frac{1}{n}\hat{H}^{\epsilon}_{\max}(A^n\,|\,B^n)_{\rho^{\otimes n}} \leqslant \frac{1}{n}\ln\big\|\mathrm{tr}_A\big(P^{S,\delta}\big)\big\|_\infty = S(A\,|\,B)_\rho + 2\delta + \frac{1}{n}\ln(\mathrm{poly}(n))\ , \tag{E.6}$$

such that, taking the limits $n \to \infty$ and then $\delta \to 0$, we see that the smooth Rényi-zero conditional entropy asymptotically converges to the von Neumann conditional entropy in the i.i.d. regime.

We now provide a construction of the universal conditional typical subspace.

Proof of Proposition 10. Let

$$P^{S,\delta}_{A^nB^n} = \sum_{\substack{\lambda,\lambda'\,:\\ \bar{S}(\lambda)-\bar{S}(\lambda')\leqslant S+2\delta}} (\mathbb{1}_{A^n}\otimes \Pi^{\lambda'}_{B^n})\,\Pi^{\lambda}_{(AB)^n}\ , \tag{E.7}$$

where the respective projectors $\Pi^{\lambda'}_{B^n}, \Pi^{\lambda}_{(AB)^n}$ refer to Schur-Weyl decompositions of $\mathscr{H}_B^{\otimes n}$ and of $(\mathscr{H}_A \otimes \mathscr{H}_B)^{\otimes n}$, respectively, with $\lambda \in \mathrm{Young}(d_Ad_B, n)$ and $\lambda' \in \mathrm{Young}(d_B, n)$. Observe that $P^{S,\delta}_{A^nB^n}$ is a projector: Each term in the sum is a projector as a product of two commuting projectors (Lemma 3), and each term of the sum acts on a different subspace of $(\mathscr{H}_A \otimes \mathscr{H}_B)^{\otimes n}$. The projector $P^{S,\delta}_{A^nB^n}$ corresponds to the measurement of the two commuting POVMs $\big\{\Pi^{\lambda}_{(AB)^n}\big\}$ and $\big\{\Pi^{\lambda'}_{B^n}\big\}$, and testing whether or not the event $\bar{S}(\lambda) - \bar{S}(\lambda') \leqslant S + 2\delta$ is satisfied. Also by construction $P^{S,\delta}_{A^nB^n}$ is permutation-invariant.

For any ρ_{AB} with $S(A|B)_\rho \leqslant S$, the probability that the measurement of $P^{S,\delta}_{A^nB^n}$ fails on $\rho^{\otimes n}_{AB}$ can be upper bounded as follows. The passing event $\bar{S}(\lambda) - \bar{S}(\lambda') \leqslant S + 2\delta$ is implied in particular by the two events (a) $\bar{S}(\lambda) \leqslant S(AB)_\rho + \delta$ and (b) $\bar{S}(\lambda') \geqslant S(B)_\rho - \delta$ happening simultaneously, recalling that $S(AB)_\rho + S(B)_\rho = S(A|B)_\rho \leqslant S$. The probability of event (a) failing is

$$\Pr\bigl[\bar{S}(\lambda) > S(AB)_\rho + \delta\bigr] \leqslant \mathrm{poly}(n)\exp(-n\eta)\ , \tag{E.8}$$

as given by (C.15), and similarly for event (b)

$$\Pr\bigl[\bar{S}(\lambda') < S(B)_\rho - \delta\bigr] \leqslant \mathrm{poly}(n)\exp(-n\eta)\ . \tag{E.9}$$

We can use the same η in both cases by picking the lesser of the two values given by (C.15), if necessary. Note furthermore that $\eta > 0$ does not depend on ρ. Hence with this η, for any ρ_{AB} we have

$$\mathrm{tr}\bigl(P^{S,\delta}_{A^nB^n}\,\rho_{AB}^{\otimes n}\bigr) \geqslant 1 - \mathrm{poly}(n)\exp(-n\eta)\ , \tag{E.10}$$

as required.

For the second property, we use Lemma 4 to write

$$\begin{aligned}
\mathrm{tr}_{A^n}\bigl(P^{S,\delta}_{A^nB^n}\bigr) &= \sum_{\substack{\lambda,\lambda'\,:\\ \bar{S}(\lambda)-\bar{S}(\lambda')\leqslant S+2\delta}} \Pi^{\lambda'}_{B^n}\ \mathrm{tr}_{A^n}\Bigl(\Pi^{\lambda}_{(AB)^n}\Bigr)\,\Pi^{\lambda'}_{B^n} \\
&\leqslant \sum_{\substack{\lambda,\lambda'\,:\\ \bar{S}(\lambda)-\bar{S}(\lambda')\leqslant S+2\delta}} \mathrm{poly}(n)\, e^{n(\bar{S}(\lambda)-\bar{S}(\lambda'))}\mathbb{1}_{B^n} \\
&\leqslant \mathrm{poly}(n)\, e^{n(S+2\delta)}\,\mathbb{1}_{B^n}\ ,
\end{aligned} \tag{E.11}$$

recalling that there are only $\mathrm{poly}(n)$ many possible Young diagrams and hence at most so many terms in the sum. ■

We may now state our universality result for trivial Hamiltonians.

Proposition 11 (Difficult direction of our main theorem, for trivial Hamiltonians)**.** *Let $\mathcal{E}_{X\to X'}$ be any completely positive, trace-preserving map, and let $\epsilon > 0$. Then,*

$$\lim \frac{1}{n} W^{\epsilon}_{X^n\to X'^n}\bigl(\mathcal{E}^{\otimes n}_{X\to X'}\,\|\,\mathbb{1}_{X^n}, \mathbb{1}_{X'^n}\bigr) \leqslant \max_{\sigma_X}\,[S(\sigma_X) - S(\mathcal{E}(\sigma_X))]\ . \tag{E.12}$$

Proof of Proposition 11. Let $V_{X\to X'E}$ be a Stinespring dilation of $\mathcal{E}_{X\to X'}$ into an environment system $E \simeq X\otimes X'$. For any n, we need to find a suitable candidate implementation $\mathcal{T}_{X^n\to X'^n}$. Let

$$S = \max_{\sigma_X}[S(\sigma_X) - S(\mathcal{E}(\sigma_X))]\ , \tag{E.13}$$

and for any $\delta > 0$ let $P^{S,\delta}_{E^nX'^n}$ be given by Proposition 10. Now define

$$\mathcal{T}_{X^n\to X'^n}(\cdot) = \mathrm{tr}_{E^n}\Bigl(P^{S,\delta}_{E^nX'^n} V^{\otimes n}_{X\to X'E}\,(\cdot)\,(V^{\dagger}_{X\leftarrow X'E})^{\otimes n} P^{S,\delta}_{E^nX'^n}\Bigr)\ , \tag{E.14}$$

noting that $\mathcal{T}_{X^n\to X'^n}$ is trace-nonincreasing by construction. Our implementation simply applies the isometry corresponding to n copies of the Stinespring representation of $\mathcal{E}_{X\to X'}$, and then projects onto the universal conditional typical subspace of the corresponding state. Let $|\sigma\rangle_{XR_X}$ be any pure state, and define $|\rho\rangle_{X'ER_X} = V_{X\to X'E}\,|\sigma\rangle_{XR_X}$. Then

$$\mathcal{T}_{X^n\to X'^n}(\sigma_{XR_X}^{\otimes n}) = \mathrm{tr}_{E^n}\left(P^{S,\delta}_{E^nX'^n}\,\rho^{\otimes n}_{X'ER_X}\,P^{S,\delta}_{E^nX'^n}\right)\ . \tag{E.15}$$

Observe that $S(E\,|\,X')_\rho = S(EX')_\rho - S(X')_\rho = S(\sigma_X) - S(\mathcal{E}(\sigma_X)) \leqslant S$, by construction. Then Proposition 10 tells us that there exists a $\eta > 0$ independent of both ρ and n such that $\mathrm{tr}\big(P^{S,\delta}_{E^nX'^n}\,\rho^{\otimes n}_{X'E}\big) \geqslant 1 - \mathrm{poly}(n)\,\exp(-n\eta)$, and hence,

$$\langle\rho|^{\otimes n}_{X'ER_X}\,P^{S,\delta}_{(X'E)^n}\,|\rho\rangle^{\otimes n}_{X'ER_X} \geqslant 1 - \mathrm{poly}(n)\,\exp(-n\eta)\ . \tag{E.16}$$

The conditions of Proposition 7 are fulfilled, with $W_{X^n\to X'^nE^n} = P^{S,\delta}_{(X'E)^n}\,V^{\otimes n}_{X\to X'E}$. Hence

$$\frac{1}{2}\big\|\mathcal{T}_{X^n\to X'^nE^n} - \mathcal{E}^{\otimes n}_{X\to X'}\big\|_\diamond \leqslant \mathrm{poly}(n)\,\exp(-n\eta/2)\ . \tag{E.17}$$

Furthermore, by Property (iii) of Proposition 10,

$$\mathcal{T}_{X^n\to X'^n}(\mathbb{1}_{X^n}) = \mathrm{tr}_{E^n}\big(P^{S,\delta}_{E^nX'^n}\big) \leqslant \mathrm{poly}(n)\,e^{n(S+2\delta)}\,\mathbb{1}_{X'^n}\ . \tag{E.18}$$

So, for n large enough the mapping $\mathcal{T}_{X^n\to X'^n}$ is a valid optimization candidate in the definition of $W^\epsilon_{X^n\to X'^n}(\mathcal{E}^{\otimes n}_{X\to X'}\,\|\,\mathbb{1}_{X^n},\mathbb{1}_{X'^n})$. The value reached is then

$$W^\epsilon_{X^n\to X'^n}(\mathcal{E}^{\otimes n}_{X\to X'}\,\|\,\mathbb{1}_{X^n},\mathbb{1}_{X'^n}) \leqslant n(S+2\delta) + \ln(\mathrm{poly}(n))\ . \tag{E.19}$$

Because $(1/n)\ln(\mathrm{poly}(n)) \to 0$ as $n\to\infty$, we have

$$\lim_{n\to\infty}\frac{1}{n}\,W^\epsilon_{X^n\to X'^n}(\mathcal{E}^{\otimes n}_{X\to X'}\,\|\,\mathbb{1}_{X^n},\mathbb{1}_{X'^n}) \leqslant S + 2\delta\ . \tag{E.20}$$

Repeating the argument with $\delta \to 0$ proves the claim. ∎

E.3. Difficult direction, full version

Here we build upon the previous section to generalize to the case $\Gamma \neq \mathbb{1}$. First, we construct a "relative" generalization of our universal conditional typical projector, which will allow to smooth the process in the general case.

Proposition 12 (Proposition II of the main text)**.** *Let $\Gamma_{AB}, \Gamma'_B \geqslant 0$. Let $x \in \mathbb{R}$. Then, for any $\delta > 0$, there exists $\xi > 0$ and for any $n \in \mathbb{N}$ there exists an operator $M^{x,\delta}_{A^nB^n}$ such that*

(i) $M^{x,\delta}_{A^nB^n}$ is permutation-invariant;

(ii) $\big(M^{x,\delta}_{A^nB^n}\big)^\dagger\,M^{x,\delta}_{A^nB^n} \leqslant \mathbb{1}$;

(iii) For any pure state $|\rho\rangle_{ABR}$ satisfying $D(\rho_{AB}\,\|\,\Gamma_{AB}) - D(\rho_B\,\|\,\Gamma'_B) \geqslant x$, then

$$\mathrm{Re}\Big\{\langle\rho|^{\otimes n}_{ABR}\,M_{A^nB^n}\,|\rho\rangle^{\otimes n}_{ABR}\Big\} \geqslant 1 - \mathrm{poly}(n)\exp(-n\xi)\ ; \tag{E.21}$$

(iv) $\mathrm{tr}_{A^n}\Big(M^{x,\delta}_{A^nB^n}\, \Gamma_{AB}^{\otimes n}\, \big(M^{x,\delta}_{A^nB^n}\big)^\dagger \Big) \leqslant \mathrm{poly}(n)\exp(-n(x-4\delta))\, \Gamma_B'^{\otimes n}$.

Furthermore, if $[\Gamma_{AB}, \mathbb{1}_A \otimes \Gamma'_B] = 0$*, then* $M_{A^nB^n}$ *can be chosen to be a projector.*

Proof of Proposition 12. Let $\big\{R^k_{A^nB^n}\big\}$ be the POVM constructed by Proposition 5 for $H_{AB} = -\ln(\Gamma_{AB})$. Similarly, let $\big\{S^\ell_{B^n}\big\}$ be the corresponding POVM constructed in Proposition 5 for $H'_B = -\ln(\Gamma'_B)$. Also, as before, we denote by $\Pi^\lambda_{A^nB^n}$ and by $\Pi^\mu_{B^n}$ the projectors onto the Schur-Weyl blocks labeled by the Young diagrams $\lambda \in \mathrm{Young}(d_Ad_B, n)$ and $\mu \in \mathrm{Young}(d_B, n)$. Let

$$M^{x,\delta}_{A^nB^n} = \sum_{\substack{k,\ell,\lambda,\mu\,:\\ k-\bar{S}(\lambda)-\ell+\bar{S}(\mu)\geqslant x-4\delta}} S^\ell_{B^n}\, \Pi^\mu_{B^n}\, \Pi^\lambda_{A^nB^n}\, R^k_{A^nB^n}\ . \tag{E.22}$$

Note that $[S^\ell_{B^n}, \Pi^\mu_{B^n}] = 0$ because $S^\ell_{B^n}$ is permutation-invariant, and that $[\mathbb{1}_{A^n} \otimes S^\ell_{B^n}, \Pi^\lambda_{A^nB^n}] = 0$ because $\mathbb{1}_{A^n} \otimes S^\ell_{B^n}$ is permutation-invariant. Recall also that $[\mathbb{1}_{A^n} \otimes \Pi^\mu_{B^n}, \Pi^\lambda_{A^nB^n}] = 0$ for the same reason. Property (i) is fulfilled by construction. Then, we have

$$\begin{aligned}
M^{x,\delta\,\dagger}_{A^nB^n} M^{x,\delta}_{A^nB^n} &= \sum_{\substack{k,\ell,\lambda,\mu,\\ k',\ell',\lambda',\mu'\,:\\ k-\bar{S}(\lambda)-\ell+\bar{S}(\mu)\geqslant x-4\delta\\ k'-\bar{S}(\lambda')-\ell'+\bar{S}(\mu')\geqslant x-4\delta}} R^k_{A^nB^n}\, \Pi^\lambda_{A^nB^n}\, \Pi^\mu_{B^n}\, S^\ell_{B^n} S^{\ell'}_{B^n}\, \Pi^{\mu'}_{B^n}\, \Pi^{\lambda'}_{A^nB^n}\, R^{k'}_{A^nB^n}\\
&= \sum_{\substack{k,k',\ell,\lambda,\mu\,:\\ k-\bar{S}(\lambda)-\ell+\bar{S}(\mu)\geqslant x-4\delta\\ k'-\bar{S}(\lambda)-\ell+\bar{S}(\mu)\geqslant x-4\delta}} R^k_{A^nB^n}\, \big(\Pi^\lambda_{A^nB^n}\, \Pi^\mu_{B^n}\, S^\ell_{B^n}\big)\, R^{k'}_{A^nB^n}\\
&= \sum_{k,k'} R^k_{A^nB^n} \left(\sum_{\substack{\ell,\lambda,\mu\\ k-\bar{S}(\lambda)-\ell+\bar{S}(\mu)\geqslant x-4\delta\\ k'-\bar{S}(\lambda)-\ell+\bar{S}(\mu)\geqslant x-4\delta}} \Pi^\lambda_{A^nB^n}\, \Pi^\mu_{B^n}\, S^\ell_{B^n} \right) R^{k'}_{A^nB^n}\\
&\leqslant \sum_{k,k'} R^k_{A^nB^n} R^{k'}_{A^nB^n}\\
&= \sum_k R^k_{A^nB^n} = \mathbb{1}_{A^nB^n}\ ,
\end{aligned} \tag{E.23}$$

recalling that the operators $(\Pi^\lambda_{A^nB^n}, \Pi^\mu_{B^n}, S^\ell_{B^n})$ form a commuting set of projectors, and where in the third line the inner sum is taken to be the zero operator if no triplet (ℓ, λ, μ) satisfies the given constraints. This shows Property (ii).

Now consider any state $|\rho\rangle_{ABR}$, where R is any reference system, and assume that $D(\rho_{AB}\,\|\,\Gamma_{AB}) - D(\rho_B\,\|\,\Gamma'_B) \geqslant x$. Observe that

$$x \leqslant -S(\rho_{AB}) - \mathrm{tr}(\rho_{AB}\ln\Gamma_{AB}) + S(\rho_B) + \mathrm{tr}(\rho_B \ln\Gamma'_B)\ . \tag{E.24}$$

We write out

$$\langle\rho|^{\otimes n}_{ABR}\, M^{x,\delta}_{A^nB^n}\, |\rho\rangle^{\otimes n}_{ABR}$$

$$
\begin{aligned}
&= \sum_{\substack{k,\ell,\lambda,\mu\,:\\ k-\bar{S}(\lambda)-\ell+\bar{S}(\mu)\geqslant x-4\delta}} \langle\rho|_{ABR}^{\otimes n}\big(S^{\ell}_{B^n}\,\Pi^{\mu}_{B^n}\,\Pi^{\lambda}_{A^nB^n}\,R^{k}_{A^nB^n}\big)\,|\rho\rangle_{ABR}^{\otimes n}\\
&= \sum_{\substack{k,\ell,\lambda,\mu\,:\\ k\geqslant -\mathrm{tr}(\rho_{AB}\ln\Gamma_{AB})-\delta\\ \bar{S}(\lambda)\leqslant S(\rho_{AB})+\delta\\ \ell\leqslant -\mathrm{tr}(\rho_B\ln\Gamma_B)+\delta\\ \bar{S}(\mu)\geqslant S(\rho_B)-\delta}} \langle\rho|_{ABR}^{\otimes n}\big(S^{\ell}_{B^n}\,\Pi^{\mu}_{B^n}\,\Pi^{\lambda}_{A^nB^n}\,R^{k}_{A^nB^n}\big)\,|\rho\rangle_{ABR}^{\otimes n}\\
&\quad+ \sum_{\substack{k,\ell,\lambda,\mu\,:\\ k-\bar{S}(\lambda)-\ell+\bar{S}(\mu)\geqslant x-4\delta\ \text{ AND}\\ [\ k< -\mathrm{tr}(\rho_{AB}\ln\Gamma_{AB})-\delta\ \text{ OR}\\ \bar{S}(\lambda)>S(\rho_{AB})+\delta\ \text{ OR}\\ \ell> -\mathrm{tr}(\rho_B\ln\Gamma'_B)+\delta\ \text{ OR}\\ \bar{S}(\mu)<S(\rho_B)-\delta\ ]}} \langle\rho|_{ABR}^{\otimes n}\big(S^{\ell}_{B^n}\,\Pi^{\mu}_{B^n}\,\Pi^{\lambda}_{A^nB^n}\,R^{k}_{A^nB^n}\big)\,|\rho\rangle_{ABR}^{\otimes n}\\
&=: \ \blacksquare_1\ +\ \blacksquare_2\ \ .
\end{aligned}
\tag{E.25}
$$

The conditions in the first sum of the last expression indeed imply that $k-\bar{S}(\lambda)-\ell+\bar{S}(\mu)\geqslant -\mathrm{tr}(\rho_{AB}\ln\Gamma_{AB})-S(\rho_{AB})+\mathrm{tr}(\rho_B\ln\Gamma'_B)+S(\rho_B)-4\delta\geqslant x-4\delta$. Consider the first term in (E.25). Define the projectors

$$X_1=\sum_{k\geqslant -\mathrm{tr}(\rho_{AB}\ln\Gamma_{AB})-\delta} R^{k}_{A^nB^n}\ ; \qquad X_1^{\perp}=\mathbb{1}-X_1\ ; \tag{E.26a}$$

$$X_2=\sum_{\bar{S}(\lambda)\leqslant S(\rho_{AB})+\delta} \Pi^{\lambda}_{A^nB^n}\ ; \qquad X_2^{\perp}=\mathbb{1}-X_2\ ; \tag{E.26b}$$

$$X_3=\sum_{\bar{S}(\mu)\geqslant S(\rho_B)-\delta} \Pi^{\mu}_{B^n}\ ; \qquad X_3^{\perp}=\mathbb{1}-X_3\ ; \tag{E.26c}$$

$$X_4=\sum_{\ell\leqslant -\mathrm{tr}(\rho_B\ln\Gamma'_B)+\delta} S^{\ell}_{B^n}\ ; \qquad X_4^{\perp}=\mathbb{1}-X_4\ , \tag{E.26d}$$

and observe that

$$\mathrm{Re}\{\ \blacksquare_1\ \}\ =\ \mathrm{Re}\Big\{\langle\rho|_{ABR}^{\otimes n}\ \big(\ X_4\ X_3\ X_2\ X_1\ \big)\ |\rho\rangle_{ABR}^{\otimes n}\Big\}\ . \tag{E.27}$$

Thanks to Proposition 5, we have $\|\ X_1^{\perp}\ |\rho\rangle_{ABR}^{\otimes n}\|\leqslant 2\exp(-n\eta/2)$, recalling that $\|P|\psi\rangle\|=\sqrt{\mathrm{tr}(P\psi)}$, and hence

$$
\begin{aligned}
&\mathrm{Re}\Big\{\langle\rho|_{ABR}^{\otimes n}\ X_4\ X_3\ X_2\ X_1\ |\rho\rangle_{ABR}^{\otimes n}\Big\} &\text{(E.28)}\\
&\quad=\mathrm{Re}\Big\{\langle\rho|_{ABR}^{\otimes n}\ X_4\ X_3\ X_2\ |\rho\rangle_{ABR}^{\otimes n}\Big\}-\mathrm{Re}\Big\{\langle\rho|_{ABR}^{\otimes n}\ X_4\ X_3\ X_2\ X_1^{\perp}\ |\rho\rangle_{ABR}^{\otimes n}\Big\} &\text{(E.29)}\\
&\quad\geqslant\mathrm{Re}\Big\{\langle\rho|_{ABR}^{\otimes n}\ X_4\ X_3\ X_2\ |\rho\rangle_{ABR}^{\otimes n}\Big\}-2\exp(-n\eta/2)\ , &\text{(E.30)}
\end{aligned}
$$

using Cauchy-Schwarz to assert that $\mathrm{Re}(\langle\chi|\psi\rangle)\leqslant|\langle\chi|\psi\rangle|\leqslant\||\chi\rangle\|\,\||\psi\rangle\|$. Similarly, using (C.15), we have $\|\ X_2^{\perp}\ |\rho\rangle_{ABR}^{\otimes n}\|\leqslant\mathrm{poly}(n)\exp(-n\eta/2)$. Also, $\|\ X_3^{\perp}\ |\rho\rangle_{ABR}^{\otimes n}\|\leqslant\mathrm{poly}(n)\exp(-n\eta/2)$, and $\|\ X_4^{\perp}\ |\rho\rangle_{ABR}^{\otimes n}\|\leqslant 2\exp(-n\eta/2)$, yielding

$$\mathrm{Re}\Big\{\langle\rho|_{ABR}^{\otimes n}\ X_4\ X_3\ X_2\ |\rho\rangle_{ABR}^{\otimes n}\Big\}\geqslant\mathrm{Re}\Big\{\langle\rho|_{ABR}^{\otimes n}\ X_4\ X_3\ |\rho\rangle_{ABR}^{\otimes n}\Big\}-\mathrm{poly}(n)\ \exp(-n\eta/2)\ ;$$

$$\mathrm{Re}\Big\{\langle\rho|_{ABR}^{\otimes n}\, X_4\, X_3\, |\rho\rangle_{ABR}^{\otimes n}\Big\} \geqslant \mathrm{Re}\Big\{\langle\rho|_{ABR}^{\otimes n}\, X_4\, |\rho\rangle_{ABR}^{\otimes n}\Big\} - \mathrm{poly}(n)\, \exp(-n\eta/2)\ ;$$
$$\mathrm{Re}\Big\{\langle\rho|_{ABR}^{\otimes n}\, X_4\, |\rho\rangle_{ABR}^{\otimes n}\Big\} \geqslant 1 - 2\, \exp(-n\eta/2)\ . \tag{E.31}$$

(We take all these η's to be the same, by choosing if necessary the minimum of the four possibly different η's.) Hence

$$\mathrm{Re}\{\ \blacksquare_1\ \}\ \geqslant\ 1 - \mathrm{poly}(n)\, \exp(-n\eta/2)\ . \tag{E.32}$$

Now consider the second term $\blacksquare_2$ of (E.25). We know that

$$\Big\|R^k_{A^nB^n}|\rho\rangle_{ABR}^{\otimes n}\Big\| \leqslant \exp(-n\eta/2) \qquad \text{if } k < -\,\mathrm{tr}(\rho_{AB}\ln\Gamma_{AB}) - \delta\ ; \tag{E.33a}$$
$$\Big\|\Pi^\lambda_{A^nB^n}|\rho\rangle_{ABR}^{\otimes n}\Big\| \leqslant \mathrm{poly}(n)\exp(-n\eta/2) \qquad \text{if } \bar{S}(\lambda) > S(\rho_{AB}) + \delta\ ; \tag{E.33b}$$
$$\Big\|S^\ell_{B^n}|\rho\rangle_{ABR}^{\otimes n}\Big\| \leqslant \exp(-n\eta/2) \qquad \text{if } \ell > -\,\mathrm{tr}(\rho_B\ln\Gamma'_B) + \delta\ ; \tag{E.33c}$$
$$\Big\|\Pi^\mu_{B^n}|\rho\rangle_{ABR}^{\otimes n}\Big\| \leqslant \mathrm{poly}(n)\exp(-n\eta/2) \qquad \text{if } \bar{S}(\mu) < S(\rho_B) - \delta\ , \tag{E.33d}$$

recalling that $\|P|\psi\rangle\| = \sqrt{\mathrm{tr}(P\psi)}$. So, for each term in the second sum of (E.25), we have

$$\begin{aligned}
&\Big|\langle\rho|_{ABR}^{\otimes n}\big(S^\ell_{B^n}\Pi^\mu_{B^n}\Pi^\lambda_{A^nB^n}R^k_{A^nB^n}\big)\,|\rho\rangle_{ABR}^{\otimes n}\Big| \\
&\quad = \Big|\big(\langle\rho|_{ABR}^{\otimes n}S^\ell_{B^n}\Pi^\mu_{B^n}\Pi^\lambda_{A^nB^n}\big)\big(R^k_{A^nB^n}\,|\rho\rangle_{ABR}^{\otimes n}\big)\Big| \\
&\quad \leqslant \Big\|\ R^k_{A^nB^n}\,|\rho\rangle_{ABR}^{\otimes n}\ \Big\|\cdot\Big\|\ \big(S^\ell_{B^n}\Pi^\mu_{B^n}\Pi^\lambda_{A^nB^n}\big)\,|\rho\rangle_{ABR}^{\otimes n}\ \Big\| \\
&\quad \leqslant \mathrm{poly}(n)\, \exp(-n\eta/2)\ ,
\end{aligned} \tag{E.34}$$

using the Cauchy-Schwarz inequality and because at least one of the four conditions is violated, causing at least one of the two the norms to decay exponentially (noting also that $S^\ell_{B^n}, \Pi^\mu_{B^n}, \Pi^\lambda_{A^nB^n}$ all commute). Because there are only at most $\mathrm{poly}(n)$ terms, we have

$$\begin{aligned}
|\ \blacksquare_2\ |\ \leqslant \sum_{\substack{k,\ell,\lambda,\mu\,:\\ k-\bar{S}(\lambda)-\ell+\bar{S}(\mu)\geqslant x-4\delta\ \ \mathrm{AND}\\ [\ k<-\mathrm{tr}(\sigma_X\ln\Gamma_X)-\delta\ \ \mathrm{OR}\\ \bar{S}(\lambda)>S(\sigma_X)+\delta\ \ \mathrm{OR}\\ \ell>-\mathrm{tr}(\rho_{X'}\ln\Gamma_{X'})+\delta\ \ \mathrm{OR}\\ \bar{S}(\mu)<S(\mathcal{E}(\sigma_X))-\delta\ ]}} &\Big|\langle\rho|_{ABR}^{\otimes n}\big(S^\ell_{B^n}\Pi^\mu_{B^n}\Pi^\lambda_{A^nB^n}R^k_{A^nB^n}\big)\,|\rho\rangle_{ABR}^{\otimes n}\Big| \\
&\leqslant \mathrm{poly}(n)\exp(-n\eta/2)\ .
\end{aligned} \tag{E.35}$$

Hence,

$$\begin{aligned}
\mathrm{Re}\Big\{\langle\rho|_{ABR}^{\otimes n}\, M^{x,\delta}_{A^nB^n}\,|\rho\rangle_{ABR}^{\otimes n}\Big\}\ &=\ \mathrm{Re}\{\ \blacksquare_1\ \}\ +\ \mathrm{Re}\{\ \blacksquare_2\ \} \\
&\geqslant\ \mathrm{Re}\{\ \blacksquare_1\ \}\ -\ |\ \blacksquare_2\ | \\
&\geqslant\ 1 - \mathrm{poly}(n)\exp(-n\eta/2)\ ,
\end{aligned} \tag{E.36}$$

proving Property (iii) for $\xi = \eta/2$. Note that ξ does not depend on the state $|\sigma\rangle_{XR}$.

Now we prove Property (iv). We have

$$\begin{aligned}
&\mathrm{tr}_{A^n}(M^{x,\delta}_{A^nB^n}\Gamma^{\otimes n}_{AB}(M^{x,\delta}_{A^nB^n})^\dagger) \\
&\qquad\leqslant \mathrm{poly}(n) \sum_{\substack{k,\ell,\lambda,\mu\,: \\ k-\bar{S}(\lambda)-\ell+\bar{S}(\mu)\geqslant x-4\delta}} \mathrm{tr}_{A^n}\big(S^\ell \Pi^\mu \Pi^\lambda R^k\, \Gamma^{\otimes n}\, R^k \Pi^\lambda \Pi^\mu S^\ell\big)\ , \qquad (E.37)
\end{aligned}$$

using Lemma 26 and dropping some subsystem indices for readability. Recall that, using Proposition 5 and Lemma 4,

$$R^k_{A^nB^n}\,\Gamma^{\otimes n}_{AB} \leqslant e^{-nk}\, R^k_{A^nB^n} \leqslant e^{-nk}\,\mathbb{1}_{A^nB^n}\ ; \qquad (E.38)$$

$$\Pi^\mu_{B^n}\,\mathrm{tr}_{A^n}\big(\Pi^\lambda_{A^nB^n}\big)\Pi^\mu_{B^n} \leqslant \mathrm{poly}(n)\,\exp(n(\bar{S}(\lambda)-\bar{S}(\mu)))\,\mathbb{1}_{B^n}\ ; \qquad (E.39)$$

$$S^\ell_{B^n} \leqslant e^{n\ell}\, S^\ell_{B^n}\,\Gamma'^{\otimes n}_B \leqslant e^{n\ell}\,\Gamma'^{\otimes n}_B\ , \qquad (E.40)$$

further recalling that $[R^k_{A^nB^n}, \Gamma^{\otimes n}_{AB}] = 0$ and $[S^\ell_{B^n}, \Gamma'^{\otimes n}_B] = 0$. Combining these together yields:

$$\begin{aligned}
(E.37) &\leqslant \mathrm{poly}(n) \sum_{\substack{k,\ell,\lambda,\mu\,: \\ k-\bar{S}(\lambda)-\ell+\bar{S}(\mu)\geqslant x-4\delta}} e^{-nk}\, S^\ell\, \Pi^\mu\, \mathrm{tr}_{A^n}\big(\Pi^\lambda_{A^nB^n}\big)\, \Pi^\mu\, S^\ell \\
&\leqslant \sum_{\substack{k,\ell,\lambda,\mu\,: \\ k-\bar{S}(\lambda)-\ell+\bar{S}(\mu)\geqslant x-4\delta}} \mathrm{poly}(n)\, e^{-nk+n(\bar{S}(\lambda)-\bar{S}(\mu))}\, S^\ell_{B^n} \\
&\leqslant \sum_{\substack{k,\ell,\lambda,\mu\,: \\ k-\bar{S}(\lambda)-\ell+\bar{S}(\mu)\geqslant x-4\delta}} \mathrm{poly}(n)\, e^{-n(k-\bar{S}(\lambda)+\bar{S}(\mu)-\ell)}\, \Gamma'^{\otimes n}_B \\
&\leqslant \mathrm{poly}(n)\, e^{-n(x-4\delta)}\, \Gamma'^{\otimes n}_B\ , \qquad (E.41)
\end{aligned}$$

as required. ∎

We can now finally prove the difficult direction of our main theorem. This proof generalizes the proof of Proposition 11 above, using now the more general universal quantum relative conditional typical smoothing operator.

Lemma 13. *Let $\Gamma_X, \Gamma_{X'} \geqslant 0$. Let $\mathcal{E}_{X\to X'}$ be any completely positive, trace-preserving map, and let $\epsilon > 0$. Then,*

$$\lim \frac{1}{n} W^\epsilon_{X^n\to X'^n}(\mathcal{E}^{\otimes n}_{X\to X'} \,\|\, \Gamma^{\otimes n}_X, \Gamma^{\otimes n}_{X'}) \leqslant \max_{\sigma_X}\,[D(\mathcal{E}(\sigma_X)\,\|\,\Gamma_{X'}) - D(\sigma_X\,\|\,\Gamma_X)]\ . \qquad (E.42)$$

Proof of Lemma 13. Let $V_{X\to X'E}$ be a Stinespring dilation of $\mathcal{E}_{X\to X'}$ into an environment system $E \simeq X\otimes X'$. For any n, we need to find a suitable candidate implementation $\mathcal{T}_{X^n\to X'^n}$. For any $\delta > 0$, let

$$x = \max_{\sigma_X}[D(\mathcal{E}(\sigma_X)\,\|\,\Gamma_{X'}) - D(\sigma_X\,\|\,\Gamma_X)]\ , \qquad (E.43)$$

and let $M^{x,\delta}_{E^nX'^n}$ be the operator constructed by Proposition 12, with the system E playing the role of the system A, with $V_{X\to X'E}\,\Gamma_X\,V^\dagger_{X\leftarrow X'E}$ as Γ_{AB} and with $\Gamma_{X'}$ as Γ'_B. Now define

$$\mathcal{T}_{X^n\to X'^n}(\cdot) = \mathrm{tr}_{E^n}\Big(M^{x,\delta}_{E^nX'^n} V^{\otimes n}_{X\to X'E}\,(\cdot)\,(V^\dagger_{X\leftarrow X'E})^{\otimes n}(M^{x,\delta}_{E^nX'^n})^\dagger\Big)\ , \tag{E.44}$$

noting that $\mathcal{T}_{X^n\to X'^n}$ is trace-nonincreasing by construction thanks to Property (ii) of Proposition 12.

Let $|\sigma\rangle_{XR_X}$ be any pure state, and define $|\rho\rangle_{X'ER_X} = V_{X\to X'E}\,|\sigma\rangle_{XR_X}$. Note that by construction, $D\big(\rho_{EX'}\,\big\|\,(V_{X\to X'E}\Gamma_X V^\dagger)\big) - D(\rho_{X'}\,\|\,\Gamma_{X'}) \geqslant x$. Then Property (iii) of Proposition 12 tells us that there exists a $\xi > 0$ independent of both ρ and n such that

$$\mathrm{Re}\Big\{\langle\rho|^{\otimes n}_{X'ER_X}\,M^{x,\delta}_{E^nX'^n}\,|\rho\rangle^{\otimes n}_{X'ER_X}\Big\} \geqslant 1 - \mathrm{poly}(n)\,\exp(-n\xi)\ . \tag{E.45}$$

The conditions of Proposition 7 are fulfilled, with $W_{X^n\to X'^nE^n} = M^{x,\delta}_{A^nB^n}\,V^{\otimes n}_{X\to X'E}$, thanks furthermore to Property (i) of Proposition 12. Hence

$$\frac{1}{2}\big\|\mathcal{T}_{X^n\to X'^nE^n} - \mathcal{E}^{\otimes n}_{X\to X'}\big\|_\diamond \leqslant \mathrm{poly}(n)\,\exp(-n\xi/2)\ . \tag{E.46}$$

Furthermore, by Property (iii) of Proposition 12, we have that

$$\begin{aligned}\mathcal{T}_{X^n\to X'^n}(\Gamma^{\otimes n}_X) &= \mathrm{tr}_{E^n}\big(M^{x,\delta}_{E^nX'^n}\,\big(V_{X\to X'E}\,\Gamma^{\otimes n}_X V^\dagger_{X\leftarrow X'E}\big)^{\otimes n}(M^{x,\delta}_{E^nX'^n})^\dagger\big)\\ &\leqslant \mathrm{poly}(n)\,e^{-n(x-4\delta)}\,\Gamma^{\otimes n}_{X'}\ .\end{aligned} \tag{E.47}$$

So, for n large enough the mapping $\mathcal{T}_{X^n\to X'^n}$ is a valid optimization candidate in the definition of $W^\epsilon_{X^n\to X'^n}(\mathcal{E}^{\otimes n}_{X\to X'}\,\|\,\Gamma^{\otimes n}_X,\Gamma^{\otimes n}_{X'})$. The attained value is then

$$W^\epsilon_{X^n\to X'^n}(\mathcal{E}^{\otimes n}_{X\to X'}\,\|\,\Gamma^{\otimes n}_X,\Gamma^{\otimes n}_{X'}) \leqslant -n(x-4\delta) + \ln(\mathrm{poly}(n))\ . \tag{E.48}$$

Because $(1/n)\ln(\mathrm{poly}(n)) \to 0$ as $n\to\infty$, we have

$$\lim_{n\to\infty}\frac{1}{n}\,W^\epsilon_{X^n\to X'^n}(\mathcal{E}^{\otimes n}_{X\to X'}\,\|\,\mathbb{1}_{X^n},\mathbb{1}_{X'^n}) \leqslant -x + 4\delta\ . \tag{E.49}$$

Repeating the argument with $\delta\to 0$ proves the claim. ∎

Appendix F: New proof of the asymptotic equiparition property of the coherent relative entropy

Here we have a new proof of the AEP for the coherent relative entropy, with an explicit expression of a smoothing process which does the job. This isn't directly useful for our universality result, but it uses similar ideas and gives some intuition about the AEP of the coherent relative entropy. Furthermore, we get an explicit process that is near-optimal in the definition of the coherent relative entropy for an i.i.d. process matrix.

Recall the definition of the relative typical subspace [20, 23]:

Proposition 14 (Relative typical projector [20, 23])**.** *Let $\rho, \tau \geqslant 0$ be operators on a finite dimensional Hilbert space $\mathscr{H}$ with $\operatorname{tr}(\rho) = 1$, and let $\delta > 0$. There exists a constant $\eta > 0$, and for all n there exists a projector $\Pi^{n,\delta}_{\rho|\tau}$ such that the following conditions hold:*

$$\big[\Pi^{n,\delta}_{\rho|\tau}, \tau^{\otimes n}\big] = 0\ ; \tag{F.1a}$$

$$e^{-n(M(\rho\,\|\,\tau)+\delta)}\,\Pi^{n,\delta}_{\rho|\tau} \ \leqslant\ \Pi^{n,\delta}_{\rho|\tau}\,\tau^{\otimes n}\,\Pi^{n,\delta}_{\rho|\tau} \ \leqslant\ e^{-n(M(\rho\,\|\,\tau)-\delta)}\,\Pi^{n,\delta}_{\rho|\tau}\ ; \tag{F.1b}$$

$$\operatorname{tr}\big(\Pi^{n,\delta}_{\rho|\tau}\,\rho^{\otimes n}\big) \geqslant 1 - 2e^{-n\eta}\ , \tag{F.1c}$$

where we have defined

$$M(\rho\,\|\,\tau) := -\operatorname{tr}(\rho\ln\tau)\ . \tag{F.2}$$

The usual (weakly) typical projector for a state ρ is obtained by choosing $\tau = \rho$:

$$\Pi^{n,\delta}_{\rho} = \Pi^{n,\delta}_{\rho|\rho}\ . \tag{F.3}$$

The construction of the relative typical projector, as well as the proof of properties (F.1a) and (F.1b) are presented in refs. [20, 23]. Here we show property (F.1c).

Proof of Proposition 14. The construction of refs. [20, 23] satisfies properties (F.1a) and (F.1b); it remains to prove (F.1c). Consider the quantity $\operatorname{tr}\big((\mathbb{1} - \Pi^{n,\delta}_{\rho|\tau})\,\rho^{\otimes n}\big)$, and note that it corresponds to the probability that a sequence of measurements of copies ρ of the observable $-\ln(\tau)$ ensemble averages to a quantity that is δ-far from $M(\rho\,\|\,\tau) = -\operatorname{tr}(\rho\ln\tau)$. Let Z_j for $j = 1, \ldots, n$ be random variables where Z_j is the outcome of the measurement of $-\ln(\tau)$ on the j-th system of $\rho^{\otimes n}$. Then using Hoeffding's inequality we find

$$\operatorname{tr}\big((\mathbb{1} - \Pi^{n,\delta}_{\rho|\tau})\,\rho^{\otimes n}\big) = \Pr\left[\left|\frac{1}{n}\sum Z_j - \big(-\operatorname{tr}(\rho\ln\tau)\big)\right| > \delta\right] \leqslant 2\exp(-n\eta)\ , \tag{F.4}$$

for some $\eta \geqslant \delta^2/\|{-\ln\tau}\|_\infty$, noting that the difference between the maximum and minimum eigenvalues of $-\ln\tau$ is upper bounded by $2\|{-\ln\tau}\|_\infty$. ■

Now we may present the new proof of the asymptotic equipartition property of the coherent relative entropy.

Proposition 15. *Let $\Gamma_X, \Gamma_{X'} \geqslant 0$, let $R_X \simeq X$ and let $|\sigma\rangle_{X:R_X}$ be any state. Write $\rho_{X'R_X} = \mathcal{E}(\sigma_{XR_X}) = \rho_{R_X}^{1/2}\,E_{X'R_X}\,\rho_{R_X}^{1/2}$ writing $E_{X'R_X} = \mathcal{E}(\Phi_{X:R_X})$. Let $\Gamma_X, \Gamma_{X'} \geqslant 0$. For any $\delta > 0$, and for any n, let*

$$S_{X'^n} = \Pi^{n,\delta}_{\rho_{X'}|\Gamma_{X'}^{-1}}\ ;\quad Q_{X'^n} = \Pi^{n,\delta}_{\rho_{X'}}\ ;\quad P_{X^n} = \Pi^{n,\delta}_{\sigma_X|\Gamma_{R_X}}\ ,\quad R_{X^n} = \Pi^{n,\delta}_{\sigma_X}\ . \tag{F.5}$$

Then the completely positive map

$$\mathcal{T}_{X^n\to X'^n}(\cdot) = S_{X'^n}\,Q_{X'^n}\,\mathcal{E}^{\otimes n}_{X\to X'}\big(R_{X^n}\,P_{X^n}\,(\cdot)\,P_{X^n}\,R_{X^n}\big)\,Q_{X'^n}\,S_{X'^n}\ , \tag{F.6}$$

is trace-nonincreasing and satisfies

$$\mathcal{T}_{X^n\to X'^n}(\Gamma_X^{\otimes n}) \leqslant e^{-n[D(\sigma_X\,\|\,\Gamma_X) - D(\rho_{X'}\,\|\,\Gamma_{X'}) - 4\delta]}\,\Gamma_{X'}^{\otimes n}\ ; \tag{F.7a}$$

$$P\big(\mathcal{T}_{X^n\to X'^n}(\sigma_{XR_X}^{\otimes n}), \rho_{X'R_X}^{\otimes n}\big) \leqslant 4e^{-n\eta'} \ , \tag{F.7b}$$

for some $\eta' > 0$. This implies that for any $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \hat{D}^{\epsilon}_{X^n\to X'^n}\big(\rho_{X'R_X}^{\otimes n} \,\big\|\, \Gamma_X^{\otimes n}, \Gamma_{X'}^{\otimes n}\big) \geqslant D(\sigma_X \,\|\, \Gamma_X) - D(\rho_{X'} \,\|\, \Gamma_{X'}) \ . \tag{F.8}$$

Proof of Proposition 15. It will prove convenient to work in the purified space, so let $E \simeq X' \otimes R_X$, and let $|E\rangle_{X'R_XE}$ be a purification of $E_{X'R_X}$. Let $V_{X\to X'E}$ be the corresponding isometry which satisfies $|E\rangle_{X'R_XE} = V_{X\to X'E}\,|\Phi\rangle_{X:R_X}$; this isometry is just a Stinespring dilation of $\mathcal{E}$. Now define

$$|T\rangle_{X'^n R_X^n E^n} = S_{X'^n}\, Q_{X'^n}\, P_{R_X^n}\, R_{R_X^n}\, |E\rangle_{X'R_XE}^{\otimes n} \ , \tag{F.9}$$

where $P_{R_X^n} = t_{X^n\to R_X^n}(P_{X^n})$ and $R_{R_X^n} = t_{X^n\to R_X^n}(R_{X^n})$. We begin by showing (F.7a). Writing $\Gamma_{R_X} = t_{X\to R_X}(\Gamma_X)$, we have

$$\begin{aligned}
&(\Gamma_{X'}^{-1/2})^{\otimes n}\ \mathrm{tr}_{R_X^n}\big[T_{X'^n R_X^n}\, \Gamma_{R_X}^{\otimes n}\big]\, (\Gamma_{X'}^{-1/2})^{\otimes n} \\
&\quad = (\Gamma_{X'}^{-1/2})^{\otimes n}\, S_{X'^n} Q_{X'^n}\ \mathrm{tr}_{R_X^n}\big[R_{R_X^n}\, E_{X'R_X}^{\otimes n}\, R_{R_X^n}\, (P_{R_X^n}\, \Gamma_{R_X}^{\otimes n}\, P_{R_X^n})\big]\, Q_{X'^n} S_{X'^n}\, (\Gamma_{X'}^{-1/2})^{\otimes n} \\
&\quad \leqslant e^{-n(M(\sigma_X \,\|\, \Gamma_X)-\delta)}\, (\Gamma_{X'}^{-1/2})^{\otimes n}\, S_{X'^n} Q_{X'^n}\ \mathrm{tr}_{R_X^n}\big[R_{R_X^n}\, E_{X'R_X}^{\otimes n}\, R_{R_X^n}\big]\, Q_{X'^n} S_{X'^n}\, (\Gamma_{X'}^{-1/2})^{\otimes n} \ ,
\end{aligned} \tag{F.10}$$

recalling that $P_{R_X^n}\, \Gamma_{R_X}^{\otimes n}\, P_{R_X^n} \leqslant e^{-n(M(\sigma_X \,\|\, \Gamma_X)-\delta)}\, \mathbb{1}_{R_X^n}$. Now define $R_{X'^n E^n}$ as the dual projector of $R_{R_X^n}$ with respect to $|E\rangle_{X'R_XE}^{\otimes n}$: Indeed, we have $|E\rangle_{X'R_XE} = V_{X\to X'E}|\Phi\rangle_{X:R_X}$; we may thus define $R_{X'^n E^n} = (V_{X\to X'E})^{\otimes n}\, R_{X^n}\, (V^\dagger)^{\otimes n}$ in such a way that $R_{X'^n E^n}\, |E\rangle_{X'R_XE}^{\otimes n} = V_{X\to X'E}^{\otimes n}\, \big(R_{X^n} \otimes \mathbb{1}_{R_X^n}\big)\, |\Phi\rangle_{X:R_X} = R_{R_X^n}\, |E\rangle_{X'R_XE}^{\otimes n}$. Then compute

$$\begin{aligned}
&Q_{X'^n}\ \mathrm{tr}_{R_X^n}\big[R_{R_X^n}\, E_{X'R_X}^{\otimes n}\, R_{R_X^n}\big]\, Q_{X'^n} \\
&\quad = Q_{X'^n}\ \mathrm{tr}_{R_X^n E^n}\big[R_{X'^n E^n}\, E_{X'R_XE}^{\otimes n}\, R_{X'^n E^n}\big]\, Q_{X'^n} \\
&\quad = Q_{X'^n}\ \mathrm{tr}_{E^n}\big[R_{X'^n E^n}\, E_{X'E}^{\otimes n}\, R_{X'^n E^n}\big]\, Q_{X'^n} \\
&\quad \leqslant Q_{X'^n}\ \mathrm{tr}_{E^n}\big[R_{X'^n E^n}\big]\, Q_{X'^n} \\
&\quad \leqslant e^{n(S(\rho_{X'E})+\delta)} Q_{X'^n}\ \mathrm{tr}_{E^n}\big[R_{X'^n E^n}\, \rho_{X'E}^{\otimes n} R_{X'^n E^n}\big]\, Q_{X'^n} \\
&\quad \leqslant e^{n(S(\rho_{X'E})+\delta)} Q_{X'^n}\ \mathrm{tr}_{E^n}\big[\rho_{X'E}^{\otimes n}\big]\, Q_{X'^n} \\
&\quad = e^{n(S(\rho_{X'E})+\delta)} Q_{X'^n}\, \rho_{X'}^{\otimes n}\, Q_{X'^n} \\
&\quad \leqslant e^{n(S(\rho_{X'E})-S(\rho_{X'})+2\delta)} Q_{X'^n} \\
&\quad \leqslant e^{n(S(\sigma_X)-S(\rho_{X'})+2\delta)} \mathbb{1}_{X'^n} \ .
\end{aligned} \tag{F.11}$$

where we have used the fact that $E_{X'E} \leqslant \mathbb{1}_{X'E}$ (since $E_{R_X} \leqslant \mathbb{1}_{R_X}$), the usual properties of the typical projectors, the fact that $[R_{X'^n E^n}, \rho_{X'E}^{\otimes n}] = 0$, as well as the fact that $S(\rho_{X'E}) = S(\rho_{R_X}) = S(\sigma_X)$ because $\rho_{X'R_XE}$ is a pure state. We may then return to

$$\begin{aligned}
\text{(F.10)} &\leqslant e^{-n(M(\sigma_X \,\|\, \Gamma_X) - S(\rho_{X'E}) + S(\rho_{X'}) - 3\delta)}\ S_{X'^n}\, (\Gamma_{X'}^{-1})^{\otimes n}\, S_{X'^n} \\
&\leqslant e^{-n(M(\sigma_X \,\|\, \Gamma_X) + M(\rho_{X'} \,\|\, \Gamma_{X'}^{-1}) - S(\rho_{X'E}) + S(\rho_{X'}) - 4\delta)}\ \mathbb{1} \ ,
\end{aligned} \tag{F.12}$$

recalling that $S_{X'^n}$ and $(\Gamma_{X'}^{-1})^{\otimes n}$ commute. A simple calculation then yields

$$\begin{aligned} &M(\sigma_X \,\|\, \Gamma_X) + M(\rho_{X'} \,\|\, \Gamma_{X'}^{-1}) - S(\sigma_X) + S(\rho_{X'}) \\ &\qquad = -\operatorname{tr}(\sigma_X \ln \Gamma_X) + \operatorname{tr}(\rho_{X'} \ln \Gamma_{X'}) + \operatorname{tr}(\sigma_X \ln \sigma_X) - \operatorname{tr}(\rho_{X'} \ln \rho_{X'}) \\ &\qquad = D(\sigma_X \,\|\, \Gamma_X) - D(\rho_{X'} \,\|\, \Gamma_{X'}) \ , \end{aligned} \tag{F.13}$$

which proves (F.7a). Now we go for (F.7b). Let $\eta_R, \eta_P, \eta_Q, \eta_S > 0$ be the corresponding parameters provided by Proposition 14 for R_{X^n}, P_{X^n}, $Q_{X'^n}$, and $S_{X'^n}$, and let $\eta = \min(\eta_R, \eta_P, \eta_Q, \eta_S)$. Then we may compute

$$\begin{aligned} &\operatorname{Re}\Big\{ \langle\rho|_{X'ER_X}^{\otimes n} \big(\sigma_{R_X}^{1/2}\big)^{\otimes n} |T\rangle_{X'^n R_X^n E^n} \Big\} \\ &\qquad = \operatorname{Re}\Big\{ \langle\rho|_{X'ER_X}^{\otimes n} S_{X'^n}\, Q_{X'^n}\, R_{X'^n E^n}\, P_{X'^n E^n}\, |\rho\rangle_{X'R_XE}^{\otimes n} \Big\} \ , \end{aligned} \tag{F.14}$$

where analogously to $R_{X'^n E^n}$ we define $P_{X'^n E^n} = V^{\otimes n} P_{X^n} (V^\dagger)^{\otimes n}$, and noting that $\big(\sigma_{R_X}^{1/2}\big)^{\otimes n} |E\rangle_{X'R_XE}^{\otimes n} = |\rho\rangle_{X'R_XE}^{\otimes n}$. Compute

$$\begin{aligned} \big\|(\mathbb{1} - P_{X'^n E^n})|\rho\rangle_{X'R_XE}^{\otimes n}\big\|^2 &= \langle\rho|_{X'R_XE}^{\otimes n} (\mathbb{1} - P_{X'^n E^n})\, |\rho\rangle_{X'R_XE}^{\otimes n} \\ &= \operatorname{tr}\big(\sigma_X\, (\mathbb{1} - P_{X^n})\big) \leqslant 2\exp(-n\eta) \ ; \end{aligned} \tag{F.15a}$$

$$\big\|(\mathbb{1} - R_{X'^n E^n})|\rho\rangle_{X'R_XE}^{\otimes n}\big\|^2 = \operatorname{tr}\big(\sigma_X\, (\mathbb{1} - R_{X^n})\big) \leqslant 2\exp(-n\eta) \ ; \tag{F.15b}$$

$$\big\|(\mathbb{1} - Q_{X'^n})|\rho\rangle_{X'R_XE}^{\otimes n}\big\|^2 = \operatorname{tr}\big(\rho_{X'}\, (\mathbb{1} - Q_{X'^n})\big) \leqslant 2\exp(-n\eta) \ ; \tag{F.15c}$$

$$\big\|(\mathbb{1} - S_{X'^n})|\rho\rangle_{X'R_XE}^{\otimes n}\big\|^2 = \operatorname{tr}\big(\rho_{X'}\, (\mathbb{1} - S_{X'^n})\big) \leqslant 2\exp(-n\eta) \ ; \tag{F.15d}$$

exploiting each time property (F.1c) of the corresponding relative typical projector. Now we use the fact that for any states $|\psi\rangle, |\psi'\rangle$, and for any $0 \leqslant X \leqslant \mathbb{1}$, we have $\operatorname{Re}\{\langle\psi'|\,X\,|\psi\rangle\} = \operatorname{Re}\{\langle\psi'|\psi\rangle\} - \operatorname{Re}\{\langle\psi'|\,(\mathbb{1}-X)\,|\psi\rangle\} \geqslant \operatorname{Re}\{\langle\psi'|\psi\rangle\} - \big\||\psi'\rangle\big\|\big\|(\mathbb{1}-X)|\psi\rangle\big\|$, where the last inequality holds by Cauchy-Schwarz. Then

$$\begin{aligned} \text{(F.14)} &\geqslant \operatorname{Re}\Big\{ \langle\rho|_{X'ER_X}^{\otimes n} S_{X'^n}\, Q_{X'^n}\, R_{X'^n E^n}\, |\rho\rangle_{X'R_XE}^{\otimes n} \Big\} - \sqrt{2}\exp(-n\eta/2) \\ &\geqslant \operatorname{Re}\Big\{ \langle\rho|_{X'ER_X}^{\otimes n} S_{X'^n}\, Q_{X'^n}\, |\rho\rangle_{X'R_XE}^{\otimes n} \Big\} - 2\sqrt{2}\exp(-n\eta/2) \\ &\geqslant \ldots \\ &\geqslant 1 - 4\sqrt{2}\exp(-n\eta/2) \ . \end{aligned} \tag{F.16}$$

Since $|\rho\rangle_{X'R_XE}^{\otimes n}$ is a purification of $\rho_{X'R_X}^{\otimes n}$, and $\big(\sigma_{R_X}^{1/2}\big)^{\otimes n} |T\rangle_{X'^n R_X^n E^n}$ is a purification of $\mathcal{T}_{X^n \to X'^n}(\sigma_{XR_X}^{\otimes n})$, we have

$$\begin{aligned} F\big(\rho_{X'R_X}^{\otimes n}, \mathcal{T}_{X^n \to X'^n}(\sigma_{XR_X}^{\otimes n})\big) &\geqslant \big|\langle\rho|_{X'ER_X}^{\otimes n} \big(\sigma_{R_X}^{1/2}\big)^{\otimes n} |T\rangle_{X'^n R_X^n E^n}\big| \\ &\geqslant \operatorname{Re}\big\{\langle\rho|_{X'ER_X}^{\otimes n} \big(\sigma_{R_X}^{1/2}\big)^{\otimes n} |T\rangle_{X'^n R_X^n E^n}\big\} \\ &\geqslant 1 - 4\sqrt{2}\exp(-n\eta/2) \ . \end{aligned} \tag{F.17}$$

Hence, and since $\bar{F}(\cdot,\cdot) \geqslant F(\cdot,\cdot)$, we have

$$P\big(\rho_{X'R_X}^{\otimes n}, \mathcal{T}_{X^n\to X'^n}(\sigma_{XR_X}^{\otimes n})\big) \leqslant \sqrt{1-(1-4\sqrt{2}\exp(-n\eta/2))^2} \leqslant \sqrt{8\sqrt{2}}\,\exp(-n\eta/4)\ , \tag{F.18}$$

using the fact that $\sqrt{1-(1-x)^2} \leqslant \sqrt{2x}$, thus proving (F.7b) noting that $(8\sqrt{2})^{1/2} \leqslant 4$. ■

Appendix G: Optimal, universal thermal operations-based implementation of any i.i.d. time-covariant channel

Here we present the proof of Theorem IV. Here again, the equality can be split into easy and a difficult directions. The easy direction follows from the Gibbs-preserving maps case, because any thermal operations implementation is also a Gibbs-preserving maps implementation. Hence, there is only the difficult direction to prove.

G.1. Building blocks for the universal protocol

We first reformulate the ideas of the convex-split lemma, the position-based decoding, and the catalytic decoupling schemes [25–29, 38] in such a way that will be most useful for our later thermodynamic application. The underlying ideas of this proposition are exactly the same as, e.g., in Ref. [26]; yet our technical statement differs in some aspects, and we provide a proof for completeness. The setting is depicted in Fig. 3.

Proposition 16 (Conditional erasure using position-based decoding)**.** *Consider two systems S, M. Fix an integer m, and let J be a large register of dimension at least $\lceil\log_2(m)\rceil + 1$ qubits, and choose a fixed basis $\{|j\rangle_J\}$. Let γ_S be any state. Let $\mathcal{S}_{SM}$ be an arbitrary set of quantum states on $S\otimes M$. Let P_{SM} be a Hermitian operator satisfying $0 \leqslant P_{SM} \leqslant \mathbb{1}$, and assume that there exists $\kappa, \kappa' \geqslant 0$ such that for all $\rho_{SM} \in \mathcal{S}_{SM}$,*

$$\mathrm{tr}\big(P_{SM}\,\rho_{SM}\big) \geqslant 1-\kappa\ ; \tag{G.1a}$$

$$\mathrm{tr}(P_{SM}\,(\gamma_S\otimes\rho_M)) \leqslant \frac{\kappa'}{m}\ . \tag{G.1b}$$

Let $A = A_1\otimes\cdots\otimes A_m$ be a monstrous collection of ancilla systems with each $A_j \simeq S$, and let $A' = A'_1\otimes\cdots\otimes A'_m$ be a copy of the full monstrous collection of ancilla systems. We write a purification of γ_{A_j} on A'_j as $|\gamma\rangle_{A_jA'_j} = \gamma_{A_j}^{1/2}\,|\Phi\rangle_{A_j:A'_j}$. Then there exists a unitary operator $W^{(m)}_{SMAJ\to SMAJ}$ satisfying the following property. For any reference system R, for any pure

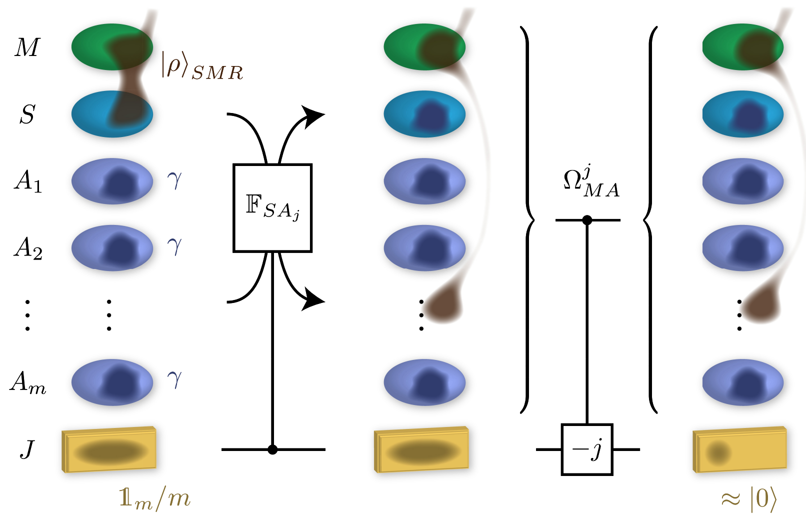

FIG. 3: Construction of the thermal operation for conditional erasure using position-based decoding [26]. A system S is to be reset to the thermal state γ_S, while exploiting the side information stored in a memory M and extracting as much work as possible. The state ρ_{MR} must be preserved, for any purifying reference system R. We bring in m ancillas $A = A_1 \ldots A_m$, where each ancilla is a copy of S, in their thermal state $\gamma_{A_j} = \gamma_S$. We bring in also a classical register J of size m, with a trivial Hamiltonian $H_J = 0$, and initialized in a maximally mixed state. Our protocol then proceeds as follows: We coherently apply the swap unitary $\mathbb{F}_{SA_j}$ between S and A_j, conditioned on the value stored in J. Then, if m is not too large, it turns out that there exists a POVM $\{\Omega^j_{MA}\}$ acting only on $MA_1 \ldots A_m$ that can infer the value j stored in the register J, up to a small error. More precisely, the maximal value of m is given by how distinguishable ρ_{SM} is from $\gamma_S \otimes \rho_M$, as quantified by a relative entropy measure. Hence we may coherently reset the J register to the zero state by conditioning on this outcome, while tolerating a small error. Because the J register is reset to a pure state, this is interpreted as extracting $kT \ln(m)$ work. In the i.i.d. regime, each of the above systems are in fact n-tensor copies. We construct in this regime a universal protocol by finding a universal POVM $\{\Omega^j_{M^n A^n}\}$ that does the job for any i.i.d. input state, by exploiting tools from Schur-Weyl duality.

tripartite state $|\rho\rangle_{SMR}$ with $\rho_{SM} \in \mathcal{S}_{SM}$, and for any $|j\rangle_J$ with $1 \leqslant j \leqslant m$, we have

$$\mathrm{Re}\Big\{ \big(\langle\hat{\tau}^j|_{RMSAA'} \otimes \langle 0|_J\big)\, W^{(m)}_{SMAJ}\, \big(|\rho\rangle_{RMS} \otimes |\gamma\rangle^{\otimes m}_{A.A'} \otimes |j\rangle_J\big) \Big\} \geqslant 1 - (2\kappa + 4\kappa')\ , \tag{G.2}$$

where we have defined

$$|\hat{\tau}^j\rangle_{RMSAA'} = |\rho\rangle_{RMA_j} \otimes |\gamma\rangle_{SA'_j} \otimes [|\gamma\rangle^{\otimes(m-1)}]_{AA'\setminus A_jA'_j}\ , \tag{G.3}$$

where by the notation $AA' \setminus A_jA'_j$ we refer to all AA' systems except $A_jA'_j$.

Furthermore, for any observables H_S, H_M such that $[P_{SM}, H_S + H_M] = 0$, the unitary $W^{(m)}_{SMAJ}$ may be chosen such that $[H_S + H_M + \sum H_{A_j}, W^{(m)}_{SMAJ}] = 0$, where $H_{A_j} = H_S$.

Intuitively, we "absorb" the initial randomness present in the register J, e.g. given to us by the environment in a mixed state, and return it in a pure state: This is work extraction. (We can return J to its initial state by running a sequence of Szilárd engines.)

Proof of Proposition 16. The operator W is defined in two steps. The first operation simply consists on conditionally swapping S with A_j, depending on the value stored in J. Then, we infer again from MA which j we swapped S with, in order to coherently reset the register J back to the zero state (approximately).

We define the first unitary operation as $W^{(1)}$, acting on systems SAJ:

$$W^{(1)}_{SAJ} = \sum_j \mathbb{F}_{SA_j} \otimes |j\rangle\langle j|_J \ , \tag{G.4}$$

where $\mathbb{F}_{SA_j}$ denotes the swap operator between the two designated systems. Observe that $W^{(1)}$ maps ρ onto $\hat{\tau}^j$ according to

$$\begin{aligned} W^{(1)}_{SQJ}\Big(|\rho\rangle_{RMS} \otimes |\gamma\rangle^{\otimes m}_{A.A'_.} \otimes |j\rangle_J\Big) \\ &= \quad |\rho\rangle_{RMA_j} \otimes |\gamma\rangle_{SA'_j} \otimes [|\gamma\rangle^{\otimes(n-1)}]_{AA'\setminus A_j A'_j} \otimes |j\rangle_J \\ &= \quad |\hat{\tau}^j\rangle_{SRMAA'} \otimes |j\rangle_J \ . \end{aligned} \tag{G.5}$$

The second step is tricky. We need to infer from the systems MA alone which j was stored in J. Fortunately the answer is provided in the form of position-based decoding [26], using a pretty good measurement. Define

$$\Lambda^j_{MA} = P_{MA_j} \otimes \mathbb{1}_{A\setminus A_j} \ , \tag{G.6}$$

such that $\{\Lambda^j_{MA}\}$ is a set of positive operators. We can form a POVM $\{\Omega^j_{MA}\}$ by normalizing the Λ^j's as follows:

$$\Omega^j_{MA} = \Lambda^{-1/2}_{MA}\, \Lambda^j_{MA}\, \Lambda^{-1/2}_{MA} \ ; \qquad\qquad \Lambda_{MA} = \sum \Lambda^j_{MA} \ . \tag{G.7}$$

We would now like to lower bound $\mathrm{tr}(\Omega^j_{MA}\hat{\tau}^j_{MA})$. Following the proof of [26, Theorem 2], we first invoke the Hayashi-Nagaoka inequality [65], which states that for any operators $0 \leqslant A \leqslant \mathbb{1}$, $B \geqslant 0$, we have

$$\mathbb{1} - (A+B)^{-1/2}\, A\, (A+B)^{-1/2} \leqslant 2(\mathbb{1} - A) + 4B \ . \tag{G.8}$$

Applying this inequality with $A = \Lambda^j_{MA}$ and $B = \sum_{j'\neq j} \Lambda^{j'}_{MA}$ we obtain

$$\begin{aligned} \mathrm{tr}\Big((\mathbb{1} - \Omega^j)\hat{\tau}^j_{MA}\Big) &\leqslant 2\,\mathrm{tr}\Big(\Big(\mathbb{1} - \Lambda^j_{MA}\Big)\hat{\tau}^j_{MA}\Big) + 4\sum_{j'\neq j} \mathrm{tr}\Big(\Lambda^{j'}_{MA}\hat{\tau}^j_{MA}\Big) \\ &\leqslant 2\,\mathrm{tr}((\mathbb{1} - P_{SM})\rho_{SM}) + 4m\ \mathrm{tr}(P_{SM}(\gamma_S \otimes \rho_M)) \\ &\leqslant 2\kappa + 4\kappa' \ . \end{aligned} \tag{G.9}$$

Now let $\mathsf{SHIFT}_J(x) = \sum_x |j+x\rangle\langle j|_J$ denote the SHIFT operation on the J register, modulo

m; note that $\big(\mathsf{SHIFT}_J(x)\big)^\dagger = \mathsf{SHIFT}_J(-x)$. We define

$$W^{(2)}_{MAJ} = \left(\sum_j \Omega^j_{MA} \otimes \mathsf{SHIFT}_J(-j)\right) ; \qquad W'_{SMAJ} = W^{(2)}_{MAJ} W^{(1)}_{SAJ} , \tag{G.10}$$

and we see that $W^\dagger W \leqslant \mathbb{1}$ thanks to Proposition 31. Then

$$\begin{aligned} &W'_{SMAJ}\Big(|\rho\rangle_{RMS} \otimes |\phi\rangle^{\otimes m}_{A.A'} \otimes |j\rangle_J\Big) \\ &\qquad = \left(\sum_{j'} \Omega^{j'}_{MA} \otimes \mathsf{SHIFT}_J(-j')\right)\big(|\hat\tau^j\rangle_{SRMAA'} \otimes |j\rangle_J\big) \\ &\qquad = \sum_{j'}\Big(\Omega^{j'}_{MA}\,|\hat\tau^j\rangle_{RMSAA'}\Big) \otimes |j-j'\rangle \ . \end{aligned} \tag{G.11}$$

Thanks to Proposition 28, the operator W'_{SMAJ} can be completed to a full unitary W_{SMAJ} by using an extra qubit in the J register, and such that $\langle 0|_J W_{SMAJ} |j\rangle_J = \langle 0|_J W'_{SMAJ} |j\rangle_J$ for all $j = 1, \ldots, m$ (with the convention that $|j\rangle_J$ for $j \leqslant m$ forces the extra qubit to be in the zero state).

So, recalling (G.9),

$$\begin{aligned} &\big(\langle\hat\tau^j|_{RMSAA'} \otimes \langle 0|_J\big) W_{SMAJ}\Big(|\rho\rangle_{RMS} \otimes |\phi\rangle^{\otimes m}_{A.A'} \otimes |j\rangle_J\Big) \\ &\quad = \big(\langle\hat\tau^j|_{RMSAA'} \otimes \langle 0|_J\big) W'_{SMAJ}\Big(|\rho\rangle_{RMS} \otimes |\phi\rangle^{\otimes m}_{A.A'} \otimes |j\rangle_J\Big) \\ &\quad = \langle\hat\tau^j|\,\Omega^j_{MA}\,|\hat\tau^j\rangle_{RMSAA'} \\ &\quad \geqslant 1 - (2\kappa + 4\kappa') \ . \end{aligned} \tag{G.12}$$

To prove the last part of the claim, let H_S, H_M be observables such that $[P_{SM}, H_S + H_M] = 0$ and $[H_S, \gamma_S] = 0$. Let $H_{A_j} = H_S$ and we write $H_A = \sum_j H_{A_j}$. For all j, we have

$$[H_S + H_M + H_A, \Lambda^j_{MA}] = \big[H_S + \textstyle\sum_{j'\neq j} H_{A_{j'}}, \Lambda^j_{MA}\big] + [H_M + H_{A_j}, P_{MA_j}] = 0 \ . \tag{G.13}$$

This implies that $[H_S + H_M + H_A, \Lambda_{MA}] = 0$, and in turn $\big[H_S + H_M + H_A, \Lambda^{-1/2}_{MA}\big] = 0$, and thus also $[H_S + H_M + H_A, \Omega^j] = 0$. Hence,

$$\big[H_S + H_M + H_A, W^{(2)}_{MAJ}\big] = 0 \ . \tag{G.14}$$

Clearly, $[H_S + H_M + H_A, W^{(1)}_{SAJ}] = 0$, and hence $[H_S + H_M + H_A, W'_{SMAJ}] = 0$. Using Proposition 29 instead of Proposition 28, we may further enforce $[H_S + H_M + H_A, W_{SMAJ}] = 0$, as required. ■

G.2. New single-shot erasure protocol for fixed input state and for noninteracting system and memory

In the position-based decoding of Ref. [26], one uses the optimal distinguishing POVM for P_{SM} obtained from the hypothesis testing entropy, and we see that for a constant error we can choose $\ln(m)$ to be proportional to the hypothesis testing entropy. In fact, this gives us directly a new erasure protocol in the case of a fixed input state, and in the case where the system and memory are not interacting ($H_{SM} = H_S + H_M$):

Corollary 17 (Proposition V of the main text)**.** *Let S, M, R be quantum systems, with H_S and H_M the Hamiltonians on S and M. For any $|\rho\rangle_{SMR}$ such that $[\rho_{SM}, H_S + H_M] = 0$, let $\epsilon > 0$ and $m = \lfloor \epsilon \exp\{D_{\mathrm{h}}^{1-\epsilon}(\rho_{SM} \,\|\, \gamma_S \otimes \rho_M)\} \rfloor$. Let J be an information battery consisting of at least $\lceil \log_2(m) \rceil + 1$ qubits. Then there exists a thermal operation acting only on S, M, J that transforms $|\rho\rangle\langle\rho|_{SMR} \otimes (\mathbb{1}_m/m)_J$ into in a state that is $\sqrt{12\epsilon}$-close to $\gamma_S \otimes \rho_{MR} \otimes |0\rangle\langle 0|_J$ in purified distance.*

Observe that the associated work extracted, counted in terms of purity using nats as units ($=$ number of qubits $\times \ln(2)$), is

$$\text{work extracted, in nats} = \ln(m) \approx D_{\mathrm{h}}^{1-\epsilon}(\rho_{SM} \,\|\, \gamma_S \otimes \rho_M) + \ln(\epsilon) \ . \tag{G.15}$$

If the register S is to be returned to a pure state instead of a thermal state, then this can be done separately as a final step, and there is a fixed cost associated to this. For instance, for trivial Hamiltonians, we have that

$$\begin{aligned} \begin{array}{c}\text{total work extracted in nats,}\\ \text{also return } S \text{ to pure state}\end{array} &\approx D_{\mathrm{h}}^{1-\epsilon}\left(\rho_{SM} \,\middle\|\, \frac{\mathbb{1}_S}{d_S} \otimes \rho_M\right) - \ln(d_S) + \ln(\epsilon) \\ &\approx D_{\mathrm{h}}^{1-\epsilon}(\rho_{SM} \,\|\, \mathbb{1}_S \otimes \rho_M) + \ln(\epsilon) \ , \end{aligned} \tag{G.16}$$

and recalling that $D_{\mathrm{h}}^{1-\epsilon}(\rho_{SM} \,\|\, \mathbb{1}_S \otimes \rho_M) \approx -H_{\max}^{\epsilon}(S \,|\, M)_\rho$ [44, 47], we recover the expression in Ref. [9] up to approximation terms.

Proof of Corollary 17. Let P_{SM} be the optimal positive semidefinite operator satisfying $0 \leqslant P_{SM} \leqslant \mathbb{1}_{SM}$ given by $D_{\mathrm{h}}^{1-\epsilon}(\rho_{SM} \,\|\, \gamma_S \otimes \rho_M)$, i.e., such that

$$\operatorname{tr}(P_{SM}\, \rho_{SM}) \geqslant 1 - \epsilon \ ; \tag{G.17}$$

$$\operatorname{tr}(P_{SM}\, (\gamma_S \otimes \rho_M)) = \exp\{-D_{\mathrm{h}}^{1-\epsilon}(\rho_{SM} \,\|\, \gamma_S \otimes \rho_M)\} \ . \tag{G.18}$$

Using the fact that $m \leqslant \epsilon \exp\{D_{\mathrm{h}}^{1-\epsilon}(\rho_{SM} \,\|\, \gamma_S \otimes \rho_M)\}$, we see that

$$\operatorname{tr}(P_{SM}\, (\gamma_S \otimes \rho_M)) \leqslant \frac{\epsilon}{m} \ . \tag{G.19}$$

We may set $\kappa = \kappa' = \epsilon$, and we are in the setting of Proposition 16 (for the single-element set $\mathcal{S}_{SM} = \{\rho_{SM}\}$). Let $W_{SMAJ}^{(m)}$ the operator given by Proposition 16, suitably extended to a unitary operator by using the last qubit of the information battery J using Proposition 28. This gives the required unitary operation for our thermal operation, using the A systems from

Proposition 16 as our bath. Indeed, the overlap of the unitary $W^{(m)}_{SMAJ}$ applied to the initial state $|\rho\rangle_{SMR} \otimes |\gamma\rangle_{AA'} \otimes |j\rangle_J$ with the state $|\hat{\tau}^j\rangle_{SMRAA'} \otimes |0\rangle_J$ is given by (G.2); we then have

$$\begin{aligned} &F\biggl(\rho_{MR} \otimes |0\rangle\langle 0|_J \,,\, \mathrm{tr}_{SA}\Bigl\{W^{(m)}_{SMAJ} \Bigl(|\rho\rangle\langle\rho|_{SMR} \otimes |\phi\rangle\langle\phi|_{AA'} \otimes \Bigl(\frac{\mathbb{1}_m}{m}\Bigr)_J\Bigr) W^{(m)\,\dagger}_{SMAJ}\Bigr\}\biggr) \\ &\quad \geqslant \frac{1}{m}\sum_j F\Bigl(\rho_{MR} \otimes |0\rangle\langle 0|_J \,,\, \mathrm{tr}_{SA}\Bigl\{W^{(m)}_{SMAJ}\,(|\rho\rangle\langle\rho|_{SMR} \otimes |\phi\rangle\langle\phi|_{AA'} \otimes |j\rangle\langle j|_J)\, W^{(m)\,\dagger}_{SMAJ}\Bigr\}\Bigr) \\ &\quad \geqslant \frac{1}{m}\sum_j \mathrm{Re}\Bigl\{\bigl(\langle\hat{\tau}^j|_{SMRAA'}\langle 0|_J\bigr)\, W^{(m)}_{SMAJ}\,(|\rho\rangle_{SMR} \otimes |\phi\rangle_{AA'} \otimes |j\rangle_J)\Bigr\} \\ &\quad \geqslant 1 - 6\epsilon \;. \end{aligned} \tag{G.20}$$

Hence, transforming the expression for the fidelity to a bound on the purified distance gives that the reduced state on SMR of the state after the noisy operation has a purified distance to $\rho_{MR} \otimes |0\rangle\langle 0|_J$ that is upper bounded by $\sqrt{12\epsilon}$. The fact that S is left in a thermal state can be enforced by an additional explicit thermalization of S that can be achieved with a simple thermal operation. ■

G.3. Universal erasure protocol for trivial Hamiltonians

As a warm-up for the general case, we construct a universal implementation of the erasure process for trivial Hamiltonians. We need a projector P_{SM} that will work for a whole class of states: We exploit the i.i.d. structure of states to construct P_{SM} using Schur-Weyl block projectors.

Theorem 18 (Universal conditional erasure protocol for trivial Hamiltonians). *Let S, M be quantum systems of dimensions d_S, d_M respectively. Let n be a positive integer.*

Let $A = A_1 \otimes \cdots \otimes A_{2d_S}$ be a monstrous collection of ancilla systems with each $A_j \simeq S$, and let $A' = A'_1 \otimes \cdots \otimes A'_{2d_S}$ be a copy of the full monstrous collection of ancilla systems. We write the maximally entangled state for each pair of ancilla systems as $|\phi\rangle_{A_j A'_j} = (\sum_k |k\rangle_{A_j} |k\rangle_{A'_j})/\sqrt{d_S}$.

Let J be a register of size $2d_S$. Let $S \in [-\ln(d_S), \ln(d_S)]$ and $\delta > 0$. Let $m = \lfloor \exp\{n(\ln(d_S) - S - 3\delta)\}\rfloor$. Then there exists an operator $\tilde{W}^{(S)}_{S^n M^n A^n J}$, and there exists $\eta' > 0$, such that:

(i) We have $\tilde{W}^\dagger \tilde{W} \leqslant \mathbb{1}$;

(ii) On any reference system R, for any $|\rho\rangle_{SMR}$ such that $S(S\,|\,M)_\rho \leqslant S$, and for any $j = 1, \ldots, m$, it holds that

$$\begin{aligned} \mathrm{Re}\Bigl\{\bigl(\langle\hat{\tau}^j|_{R^n M^n S^n A^n A'^n} \otimes \langle 0|_J\bigr)\, \tilde{W}^{(S)}_{M^n S^n A^n J}\,\Bigl(|\rho\rangle^{\otimes n}_{RMS} \otimes |\phi\rangle^{\otimes mn}_{A.A'_.} \otimes |j\rangle_J\Bigr)\Bigr\} \\ \geqslant 1 - \mathrm{poly}(n)\exp(-n\eta') \;, \end{aligned} \tag{G.21}$$

where $|\hat{\tau}^j\rangle$ is given as

$$|\hat{\tau}^j\rangle_{R^n M^n S^n A^n A'^n} = |\rho\rangle^{\otimes n}_{RMA_j} \otimes |\phi\rangle^{\otimes n}_{SA'_j} \otimes |\phi\rangle^{\otimes n}_{AA' \setminus A_j A'_j} \;. \tag{G.22}$$

Proof of Theorem 18. This is in fact a relatively straightforward application of Proposition 16 over n copies of SM. We use

$$\mathcal{S}_{S^n M^n} = \left\{ \rho_{SM}^{\otimes n} \ : \ S(S \,|\, M)_\rho \leqslant S \right\} ; \tag{G.23}$$

$$m = \lfloor \exp\{ n(\ln(d_S) - S - 3\delta) \} \rfloor \ , \tag{G.24}$$

and let $P_{S^n M^n}$ given by Proposition 10 for the current values of S, δ. We then have $\kappa = \mathrm{poly}(n) \exp\{-n\eta(\delta)\}$ from Proposition 10, and also

$$\begin{aligned}\operatorname{tr}\!\left(P_{S^n M^n} \left(\frac{\mathbb{1}_S}{d_S} \otimes \rho_M^{\otimes n} \right) \right) &\leqslant \mathrm{poly}(n)\, e^{n(S+2\delta)}\, d_S^{-n}\ \operatorname{tr}(\rho_M^{\otimes n}) = \mathrm{poly}(n)\, e^{-n(\ln(d_S) - S - 2\delta)} \\ &= \frac{\mathrm{poly}(n)\, e^{-n\delta}}{m} \ , \end{aligned} \tag{G.25}$$

and thus we may take $\kappa' = \mathrm{poly}(n)\, e^{-n\delta}$. Finally, η' is given as $\eta' = \min(\delta, \eta(\delta))$. ■

Corollary 19. *For any completely positive map $\mathcal{E}_{X \to X'}$, for any n, for any $\epsilon > 0$, and for any $\delta > 0$, there exists a joint noisy operation that acts on $X^n W \to X'^n W$ with an information battery W, using a deterministic amount of work per copy that is asymptotically equal to $\max_{\sigma_X} [S(\sigma) - S(\mathcal{E}(\sigma))] + 3\delta$, such that the induced channel on $X \to X'$ is ϵ-close to $\mathcal{E}_{X \to X'}$ in diamond norm distance.*

We defer the proof to the next section, where we directly prove the general case.

G.4. Universal implementation of any covariant process, for nontrivial Hamiltonians

We can generalize the above idea pretty straightforwardly in the case of time-covariant processes, i.e., processes that commute with the time evolution. Indeed, time-covariant processes have the nice property of mapping directly to a conditional erasure problem with a system and memory that are noninteracting. We first show this formally:

Lemma 20 (Covariant processes)**.** *Let X be a quantum system with Hamiltonian H_X. Let $\mathcal{E}_{X \to X}$ be a completely positive, trace-preserving map that is covariant with respect to time evolution; i.e., for all t,*

$$\mathcal{E}_{X \to X}(e^{-iH_X t}\, (\cdot)\, e^{iH_X t}) = e^{-iH_X t}\, \mathcal{E}_{X \to X}(\cdot)\, e^{iH_X t} \ . \tag{G.26}$$

Then there exists a system E with Hamiltonian H_E including an eigenstate $|0\rangle_E$ of zero energy, as well as a unitary $V_{EX \to EX}$, such that

$$\mathcal{E}_{X \to X}(\cdot) = \operatorname{tr}_E\!\big(V\, (|0\rangle\langle 0|_E \otimes (\cdot))\, V^\dagger \big) \ , \tag{G.27}$$

as well as $V\,(H_X + H_E)\, V^\dagger = H_X + H_E$.

The idea is then to use the convex-split/position-based decoding approach with ancillas in Gibbs states instead of being maximally mixed. This directly gives us the desired result.

Theorem 21. *Let X be a quantum system, let H_X be a Hermitian operator, and let $\beta \geqslant 0$. Let $\mathcal{E}_{X\to X}$ be a completely positive, trace-preserving map, that is time-covariant: $\mathcal{E}_{X\to X}(e^{-iHt}\,(\cdot)\,e^{iHt}) = e^{-iHt}\,\mathcal{E}_{X\to X}(\cdot)\,e^{iHt}$ for all t. Let $\delta > 0$ be small enough and let n be a positive integer. Let W be a work bit with two levels $|0\rangle_W$ and $|w\rangle_W$, where w is a fixed value satisfying*

$$\frac{w}{n} = \beta^{-1} \max_{\sigma}\left\{ D\big(\mathcal{E}(\sigma)\,\big\|\,e^{-\beta H_{X^n}}\big) - D\big(\sigma\,\big\|\,e^{-\beta H_{X^n}}\big) \right\} + \eta' \ , \tag{G.28}$$

for some specific η' given in terms of δ (with $\eta' \to 0$ as $\delta \to 0$). Then there exists a bath B and a Hamiltonian H_B, as well as a unitary operation U_{X^nWB}, such that

(i) The mapping

$$\Phi_{X^n\to X^n}(\cdot) = \operatorname{tr}_B\big(\langle 0|_W\, U_{X^nWB}\, ((\cdot) \otimes |w\rangle\langle w|_W \otimes \gamma_B)\, U^\dagger\, |0\rangle_W \big) \tag{G.29}$$

is trace-preserving, where $\gamma_B = e^{-\beta H_B} / \operatorname{tr}(e^{-\beta H_B})$;

(ii) There exists $\eta'' > 0$ independent of n such that the map $\Phi_{X^n\to X^n}$ satisfies

$$\big\| \Phi_{X^n\to X^n} - \mathcal{E}_{X\to X}^{\otimes n} \big\|_\diamond \leqslant \operatorname{poly}(n) \exp(-n\eta'') \ ; \tag{G.30}$$

(iii) The unitary U_{X^nWB} commutes with the total Hamiltonian, i.e.,

$$U_{X^nWB}\,(H_{X^n} + H_W + H_B)\, U^\dagger = H_{X^n} + H_W + H_B \tag{G.31}$$

We begin by proving the lemma:

Proof of Lemma 20. Let $V'_{X\to XE}$ be any Stinespring dilation isometry of $\mathcal{E}_{X\to X}$, such that $\mathcal{E}_{X\to X}(\cdot) = \operatorname{tr}_E\big(V'_{X\to XE}\,(\cdot)\,V'^\dagger\big)$.

For the input state $|\Phi\rangle_{X:R_X}$, consider the output state $|\varphi\rangle_{XER_X}$ corresponding to first time-evolving by some time t, and then applying V':

$$|\varphi\rangle_{XER_X} = V'\, e^{-iH_X t}\, |\Phi\rangle_{X:R_X} = e^{-i\tilde{H}_{XE} t}\, V'\, |\Phi\rangle_{X:R_X} \ , \tag{G.32}$$

where we have defined $\tilde{H}_{XE} = V'\,H_X\,V'^\dagger$. On the other hand, define $|\varphi'\rangle_{XER_X} = e^{-iH_X t}\,V'\,|\Phi\rangle_{X:R_X}$. By the covariance property of $\mathcal{E}_{X\to X}$, both $|\varphi\rangle$ and $|\varphi'\rangle$ have the same reduced state on XR_X; they are hence related by some unitary $W_E^{(t)}$ on the system E which in general depends on t:

$$|\varphi\rangle_{XER_X} = W_E^{(t)}\, |\varphi'\rangle_{XER_X} \ . \tag{G.33}$$

We have

$$\operatorname{tr}_X\big[V' e^{-iH_X t} \Phi_{X:R_X} e^{iH_X t} V'^\dagger \big] = W_E^{(t)}\, \operatorname{tr}_X\big[V' \Phi_{X:R_X} V'^\dagger \big] W_E^{(t)\,\dagger} \ , \tag{G.34}$$

so $W_E^{(t)}$ must define a representation of time evolution, at least on the support of $\operatorname{tr}_X\big[V'\Phi_{X:R_X}V'^\dagger\big]$ (we don't care what happens outside of this support). Hence we may write

$W_E^{(t)} = e^{-iH_E t}$ for some Hamiltonian H_E, and from (G.33), we have for all t

$$V'_{X\to XE}\, e^{-iH_X t} = e^{-i(H_X+H_E)t}\, V'_{X\to XE}\ . \tag{G.35}$$

Expanding for infinitesimal t we obtain

$$V'_{X\to XE}\, H_X = (H_X + H_E)\, V'_{X\to XE} \tag{G.36}$$

Let $|0\rangle_E$ be an eigenvector of H_E corresponding to the eigenvalue zero; if H_E doesn't contain an eigenvector with eigenvalue equal to zero, we may trivially add a dimension to the system E to accommodate this vector. Then the operator

$$V'_{X\to XE}\langle 0|_E \tag{G.37}$$

maps each state of a subset of energy levels of XE to a corresponding energy level of same energy on XE; it may thus be completed to a fully energy-preserving unitary $V_{XE\to XE}$. More precisely, let $|j\rangle_X$ be a complete set of eigenvectors of H_X with energies h_j. Then $|\psi'_j\rangle := V'_{X\to XE}|j\rangle_X$ is an eigenvector of $H_X + H_E$ of energy h_j thanks to (G.36). We have two orthonormal sets $\{|0\rangle_E \otimes |j\rangle_X\}$ and $\big\{|\psi'_j\rangle_X\big\}$ in which the j-th vector of each set has the same energy; we can thus complete these sets into two bases $\{|\chi_i\rangle_{XE}\}$, $\{|\chi'_i\rangle_{XE}\}$ of eigenvectors of $H_X + H_E$, where the i-th element of either basis has exactly the same energy. This defines a unitary $V_{XE\to XE} = \sum_i |\chi'_i\rangle_{XE}\langle\chi_i|_{XE}$ that is an extension of $V'_{X\to XE}\langle 0|_E$, and that satisfies all the conditions of the claim. ∎

We may now prove our main theorem for thermal operations:

Proof of Theorem 21. Thanks to Lemma 20, there exists an environment system E with Hamiltonian H_E, as well as an energy-conserving unitary V_{XE} and a state $|0\rangle_E$ of zero energy such that

$$\mathcal{E}_{X\to X}(\cdot) = \operatorname{tr}_E\Big\{V_{XE}\,(|0\rangle\langle 0|_E \otimes (\cdot))\,V_{XE}^\dagger\Big\}\ . \tag{G.38}$$

Let $F_E = -\beta^{-1}\ln(Z_E)$ with $Z_E = \operatorname{tr}(e^{-\beta H_E})$. Define

$$x = \min_\sigma\big\{D(\sigma\,\|\,e^{-\beta H_X}) - D(\mathcal{E}(\sigma)\,\|\,e^{-\beta H_X})\big\}\ . \tag{G.39}$$

Writing $\rho_{XE} = V_{XE}(|0\rangle\langle 0|_E \otimes \sigma_X)V_{XE}^\dagger$, we have that $x = \min_{\sigma_X}\big\{-S(\sigma_X) + \beta\operatorname{tr}(\sigma_X H_X) + S(\rho_X) - \beta\operatorname{tr}(\rho H_X)\big\}$. Using $\operatorname{tr}(\sigma_X H_X) = \operatorname{tr}\big((|0\rangle\langle 0|_E \otimes \sigma_X)(H_X + H_E)\big) = \operatorname{tr}\big(\rho_{XE}(H_X + H_E)\big)$, we finally see that

$$x = \min_{\sigma_X}\{-S(\rho_{XE}) + S(\rho_X) + \beta\operatorname{tr}(\rho_E H_E)\}\ . \tag{G.40}$$

Also, observe that for any such ρ_{XE}, we have

$$\begin{aligned}-S(E\,|\,X)_\rho + \beta\operatorname{tr}(\rho_E H_E) &\geqslant -S(E)_\rho + \beta\operatorname{tr}(\rho_E H_E) + \ln(Z) - \ln(Z)\\ &= D(\rho_E\,\|\,\gamma_E) + \beta F_E \geqslant \beta F_E\ ,\end{aligned} \tag{G.41}$$

using the subadditivity of the von Neumann entropy and the fact that relative entropy is positive for normalized states. Hence $x \geqslant \beta F_E$.

The idea is to use a form of convex-split construction as for the trivial Hamiltonians case (Theorem 18), but now the mixed ancillas will be replaced by ancillas in Gibbs states. First, in preparation for applying Proposition 16, we need to determine a distinguishing operator $P_{E^n X^n}$ that will successfully select $\rho_{EX}^{\otimes n}$ but that will reject $\rho_X^{\otimes n} \otimes \gamma_E^{\otimes n}$ at a suitable rate.

Let $\left\{R_{E^n}^k\right\}$ be the projectors onto the eigenspaces of H_{E^n}, corresponding to eigenvalues h_k. Consider the projector

$$P_{E^n X^n} = \sum_{\substack{\lambda,\lambda',k:\\ -S(\lambda/n)+S(\lambda'/n)+\beta h_k \geqslant x-3\delta}} \Pi_{E^n X^n}^{\lambda}\, \Pi_{X^n}^{\lambda'}\, R_{E^n}^k\ , \tag{G.42}$$

noting that all inner projectors in the sum commute for all values of λ, λ', k. We have

$$\begin{aligned} 1 - \operatorname{tr}\bigl(P_{E^n X^n}\rho_{EX}^{\otimes n}\bigr) &= \sum_{\substack{\lambda,\lambda',k:\\ -S(\lambda/n)+S(\lambda'/n)+\beta h_k < x-3\delta}} \operatorname{tr}\Bigl[\Pi_{E^n X^n}^{\lambda}\, \Pi_{X^n}^{\lambda'}\, R_{E^n}^k\, \rho_{EX}^{\otimes n}\Bigr] \\ &\leqslant \sum_{\substack{\lambda,\lambda',k:\\ S(\lambda/n)>S(\rho_{XE})+\delta \quad \text{OR}\\ S(\lambda'/n)<S(\rho_X)-\delta \quad \text{OR}\\ h_k<\operatorname{tr}(\rho_E H_E)-\delta}} \operatorname{tr}\Bigl[\Pi_{E^n X^n}^{\lambda}\, \Pi_{X^n}^{\lambda'}\, R_{E^n}^k\, \rho_{EX}^{\otimes n}\Bigr]\ , \end{aligned} \tag{G.43}$$

because if all three conditions $S(\lambda/n) \leqslant S(\rho_{XE}) + \delta$, $S(\lambda'/n) \geqslant S(\rho_X) - \delta$, and $h_k \geqslant \operatorname{tr}(\rho_E H_E) - \delta$ are satisfied, then $-S(\lambda/n) + S(\lambda'/n) + \beta h_k \geqslant x - 3\delta$, and hence each term in the first sum is included in the second. Now consider the term in the sum. In the case where $S(\lambda/n) > S(\rho_{XE}) + \delta$, then there exists $\eta > 0$ such that $\operatorname{tr}\bigl(\Pi_{E^n X^n}^{\lambda}\rho_{XE}^{\otimes n}\bigr) \leqslant \operatorname{poly}(n)\exp(-n\eta)$; in the case where $S(\lambda'/n) < S(\rho_X) - \delta$, there exists $\eta > 0$ such that $\operatorname{tr}\bigl(\Pi_{X^n}^{\lambda'}\rho_X^{\otimes n}\bigr) \leqslant \operatorname{poly}(n)\exp(-n\eta)$; and in the case where $h_k < \operatorname{tr}(\rho_E H_E) - \delta$, there exists $\eta > 0$ such that $\operatorname{tr}\bigl(R_{E^n}^k\rho_E^{\otimes n}\bigr) \leqslant 2\exp(-n\eta)$. In any case, choosing the smallest of all these η, and because there are at most $\operatorname{poly}(n)$ terms in the sum, we finally have

$$1 - \operatorname{tr}\bigl(P_{E^n X^n}\,\rho_{EX}^{\otimes n}\bigr) \leqslant \operatorname{poly}(n)\exp(-n\eta)\ . \tag{G.44}$$

On the other hand, we have

$$\operatorname{tr}\bigl(P_{E^n X^n}\,\bigl(\rho_X^{\otimes n}\otimes\gamma_E^{\otimes n}\bigr)\bigr) = \sum_{\substack{\lambda,\lambda',k:\\ -S(\lambda/n)+S(\lambda'/n)+\beta h_k \geqslant x-3\delta}} \operatorname{tr}\Bigl(\Pi_{E^n X^n}^{\lambda}\, \Pi_{X^n}^{\lambda'}\, R_{E^n}^k\, \bigl(\rho_X^{\otimes n}\otimes\gamma_E^{\otimes n}\bigr)\Bigr)\ . \tag{G.45}$$

We may bound each term of the sum as

$$\begin{aligned} \operatorname{tr}\Bigl(\Pi_{E^n X^n}^{\lambda}\, \Pi_{X^n}^{\lambda'}\, R_{E^n}^k\, \bigl(\rho_X^{\otimes n}\otimes\gamma_E^{\otimes n}\bigr)\Bigr) &\leqslant e^{-n\beta(h_k-F_E)}\operatorname{tr}\Bigl(\Pi_{E^n X^n}^{\lambda}\, \Pi_{X^n}^{\lambda'}\rho_X^{\otimes n}\Bigr) \\ &\leqslant \operatorname{poly}(n)\, e^{-n(\beta(h_k-F_E)-S(\lambda/n)+S(\lambda'/n))} \\ &\leqslant \operatorname{poly}(n)\, e^{-n(x-\beta F_E-3\delta)}\ . \end{aligned} \tag{G.46}$$

using respectively the fact that $R^k_{E^n}\gamma_E^{\otimes n} \leqslant (e^{-n\beta h_k}/Z_E^n)\,\mathbb{1}_{E^n} = e^{-n\beta(h_k - F_E)}\,\mathbb{1}_{E^n}$, the fact that $\mathrm{tr}_{E^n}\big[\Pi^{\lambda}_{E^nX^n}\Pi^{\lambda'}_{X^n}\big] \leqslant \mathrm{poly}(n)\,\exp(n(S(\lambda/n) - S(\lambda'/n)))\,\mathbb{1}_{X^n}$ (cf. Lemma 4), and the condition on the sum.

Consider a mega-ancilla A^n composed of m copies $A_1^n \ldots A_m^n$ of E^n, each with the same corresponding Hamiltonian $H_{A_j^n} = H_{E^n}$. We choose

$$m = \lfloor \exp(-n\eta)\,\exp(n(x - \beta F_E - 3\delta)) \rfloor \ . \tag{G.47}$$

(In the following, if $x = \beta F_E$, then we set $m = 1$ and the protocol is trivial. So in the following we assume that $x < \beta F_E$. Furthermore we assume that δ, η are small enough such that $3\delta + \eta < (x - \beta F_E)$.) Then, with $\kappa = \mathrm{poly}(n)\,e^{-n\eta}$ and $\kappa' = \mathrm{poly}(n)\,e^{-n\eta}$, we may apply Proposition 16 to the set of states

$$\begin{aligned}\mathcal{S}_{E^nX^n} &= \big\{\rho_{EX}^{\otimes n} \ : \ -S(E\,|\,X)_\rho + \mathrm{tr}(\rho_E H_E) \geqslant x\big\} \\ &\supset \big\{\rho_{EX}^{\otimes n} \ : \ \rho_{EX} = V_{XE}(|0\rangle\langle 0|_E \otimes \sigma_X)V_{XE}^\dagger \text{ for some } \sigma_X\big\} \ .\end{aligned} \tag{G.48}$$

Let $W^{(m)}_{E^nX^nA^nJ}$ be the corresponding unitary given by Proposition 16 where J is an information battery, and where $W^{(m)}_{E^nX^nA^nJ}$ is energy-conserving. For any state $|\sigma\rangle_{XR}$ with any reference system R, let $|\rho\rangle_{EXR} = V_{XE}\,\big(|0\rangle_E \otimes |\sigma\rangle_{XR}\big)$; then from Proposition 16 it holds that for all $j = 1, \ldots, m$,

$$\begin{aligned}\mathrm{Re}\Big\{\big(\langle \hat\tau^j(\sigma)|_{R^nE^nX^nA^nA'^n} \otimes \langle 0|_J\big)\, W^{(m)}_{E^nX^nA^nJ}\,\big(|\rho\rangle^{\otimes n}_{EXR} \otimes |\gamma\rangle_{A^nA'^n} \otimes |j\rangle_J\big)\Big\} \\ \geqslant 1 - (2\kappa + 4\kappa') \geqslant 1 - \mathrm{poly}(n)\exp(-n\eta) \ ,\end{aligned} \tag{G.49}$$

where

$$\begin{aligned}|\hat\tau^j(\sigma)\rangle_{R^nE^nX^nA^nA'^n} &= \big(V_{A_jX}|0\rangle_{A_j}|\sigma\rangle_{XR}\big)^{\otimes n} \otimes |\gamma\rangle_{E^nA_j'^n} \otimes \big[|\gamma_{E^n}\rangle^{\otimes(m-1)}\big]_{A^nA'^n\setminus A_j^nA_j'^n} \\ &= |\rho\rangle^{\otimes n}_{A_jXR} \otimes |\gamma\rangle_{E^nA_j'^n} \otimes \big[|\gamma_{E^n}\rangle^{\otimes(m-1)}\big]_{A^nA'^n\setminus A_j^nA_j'^n} \ .\end{aligned} \tag{G.50}$$

Now we can start piecing up together our full processes. Suppose we start with ancillas E^n in the state $|0\rangle_E^{\otimes n}$, and an information battery J in the state $\mathbb{1}_m/m$. Define the operator

$$\bar W_{E^nX^nA^nJ} = |0\rangle\langle 0|_J W^{(m)}_{E^nX^nA^nJ}\, V_{XE}^{\otimes n} \ , \tag{G.51}$$

noting that it commutes with the total Hamiltonian $H_{E^nX^nA^nJ} = H_{E^n} + H_{X^n} + \sum H_{A_j^n}$ and that $\bar W^\dagger \bar W \leqslant \mathbb{1}$. Invoking Proposition 30, we expand this operator to a full energy-conserving unitary $\tilde W_{E^nX^nA^nJ}$ that has the property of preserving high overlaps of the form $\mathrm{Re}\big\{\langle\psi'|\bar W|\psi\rangle\big\}$: For any $|\psi\rangle, |\psi'\rangle$ satisfying $\mathrm{Re}\big\{\langle\psi'|\bar W|\psi\rangle\big\} \geqslant 1 - \mathrm{poly}(n)\exp(-n\eta)$ then we have $\mathrm{Re}\big\{\langle\psi'|\tilde W|\psi\rangle\big\} \geqslant 1 - \mathrm{poly}(n)\exp(-n\eta/4)$.

We now invoke the postselection technique and show that this unitary, when applied onto the de Finetti state, yields an output on X^n and a reference system that is exponentially close to the ideal channel applied onto the de Finetti state. We use the expression (C.21) for the de Finetti state, i.e., $|\tau\rangle_{X^n\bar R^nR'} = \sum_i \sqrt{p_i}|\phi_i\rangle^{\otimes n}_{X\bar R} \otimes |i\rangle_{R'}$. Let $|\hat\tau'^j\rangle_{E^nX^n\bar R^nR'A^nA'^n} = \sum_i \sqrt{p_i}|\hat\tau^j(\phi_i)\rangle_{E^nX^n\bar R^nA^nA'^n} \otimes |i\rangle_{R'}$, where $|\hat\tau^j(\phi_i)\rangle$ is determined from $|\phi_i\rangle_{X\bar R}$ via (G.50).

We then have for all j, and

$$
\begin{aligned}
&\mathrm{Re}\Big\{ \big(\langle\hat{\tau}'^j|_{E^nX^n\bar{R}^nR'A^nA'^n} \otimes \langle 0|_J\big)\, \bar{W}_{E^nX^nA^nJ}\, \big(|0\rangle_E^{\otimes n} \otimes |\tau\rangle_{X^n\bar{R}^nR'} \otimes |\gamma\rangle_{A^nA'^n} \otimes |j\rangle_J\big)\Big\} \\
&\qquad = \sum_i p_i \,\mathrm{Re}\Big\{ \big(\langle\hat{\tau}^j(\phi_i)|_{\bar{R}^nE^nX^nA^nA'^n} \otimes \langle 0|_J\big)\, W^{(m)}_{E^nX^nA^nJ}\, \big(|\rho_i\rangle_{EX\bar{R}}^{\otimes n} \otimes |\gamma\rangle_{A^nA'^n} \otimes |j\rangle_J\big)\Big\} \\
&\qquad \geqslant 1 - \mathrm{poly}(n)\exp(-n\eta)\ , \qquad\qquad \text{(G.52)}
\end{aligned}
$$

where $|\rho_i\rangle_{EX\bar{R}} = V_{XE}\,(|0\rangle_E \otimes |\phi_i\rangle_{X\bar{R}})$, which implies from Proposition 30 that

$$
\begin{aligned}
&\mathrm{Re}\Big\{ \big(\langle\hat{\tau}'^j|_{E^nX^n\bar{R}^nR'A^nA'^n} \otimes \langle 0|_J\big)\, \tilde{W}_{E^nX^nA^nJ}\, \big(|0\rangle_E^{\otimes n} \otimes |\tau\rangle_{X^n\bar{R}^nR'} \otimes |\gamma\rangle_{A^nA'^n} \otimes |j\rangle_J\big)\Big\} \\
&\qquad\qquad \geqslant 1 - \mathrm{poly}(n)\exp(-n\eta/4)\ , \qquad \text{(G.53)}
\end{aligned}
$$

and where we also have

$$
\begin{aligned}
\mathrm{tr}_{A^nA'^nE^n}\big[|\hat{\tau}'^j\rangle\langle\hat{\tau}'^j|\big] &= \mathrm{tr}_{A_j^n}\Big[V_{XA_j}^{\otimes n}\Big(|0\rangle\langle 0|_E^{\otimes n} \otimes |\tau\rangle\langle\tau|_{X^n\bar{R}^nR'}\Big)(V_{XA_j}^{\dagger})^{\otimes n}\Big] \\
&= \mathcal{E}_{X\to X}^{\otimes n}(|\tau\rangle\langle\tau|_{X^n\bar{R}^nR'})\ . \qquad \text{(G.54)}
\end{aligned}
$$

Hence, defining the trace-preserving mapping

$$
\Phi_{X^n\to X^n}(\cdot) = \mathrm{tr}_{E^nA^nJ}\Big(\tilde{W}_{E^nX^nA^nJ}\,\big((\cdot) \otimes |0\rangle\langle 0|_E^{\otimes n} \otimes \gamma_{A^n} \otimes (\mathbb{1}_m/m)_J\big)\,\tilde{W}^{\dagger}_{E^nX^nA^nJ}\Big) \quad \text{(G.55)}
$$

we have

$$
F(\Phi_{X^n\to X^n}(\tau_{X^n\bar{R}^nR'}), \mathcal{E}_{X\to X}^{\otimes n}(\tau_{X^n\bar{R}^nR'})) \geqslant 1 - \mathrm{poly}(n)\exp(-n\eta/4)\ , \qquad \text{(G.56)}
$$

and hence by the postselection technique

$$
\begin{aligned}
\big\|\Phi_{X^n\to X^n} - \mathcal{E}_{X\to X}^{\otimes n}\big\|_{\diamond} &\leqslant \mathrm{poly}(n)\,\big\|\Phi_{X^n\to X^n}(\tau_{X^n\bar{R}^nR'}) - \mathcal{E}_{X\to X}^{\otimes n}(\tau_{X^n\bar{R}^nR'})\big\|_1 \\
&\leqslant \mathrm{poly}(n)\exp(-n\eta/8)\ . \qquad \text{(G.57)}
\end{aligned}
$$

So far, we have established the following: Given ancillas E^n in the state $|0\rangle_E^{\otimes n}$, a megaginormous ancilla A^n in the state γ_{A^n}, and an information battery J initialized in the state $\mathbb{1}_m/m$, then there exists an energy-conserving unitary that universally implements $\mathcal{E}_{X\to X}^{\otimes n}$ up to exponentially good precision, leaving J in a pure state. It remains to actually prepare these initial states, and extract work from the J register in its final $|0\rangle_J$ state, to obtain the final work rate. Also, we will see that we can collapse all these steps into a single energy-conserving unitary without having to make explicit reference to additional work storage systems such as the J register.

Let precisely

$$
w = -\beta^{-1}\,\ln(m) - nF_E\ , \qquad \text{(G.58)}
$$

and consider a work storage system W_0 consisting of the energy levels $|w\rangle_{W_0}$, $|w + nF_E\rangle_{W_0}$, $|-\beta^{-1}\ln(d_J)\rangle_{W_0}$ and $|0\rangle_{W_0}$. From known results in thermomajorization and thermal operations, there exists an energy-conserving unitary $K^{(1)}_{E^nW_0B_1}$ with a bath B_1 such that (in the limit of a very large bath),

$$\mathrm{tr}_{B_1}\Big[K^{(1)}_{E^nW_0B_1}\big(\gamma_E^{\otimes n}\otimes|w\rangle\langle w|_{W_0}\otimes\gamma_{B_1}\big)K^{(1)\,\dagger}_{E^nW_0B_1}\Big] = |0\rangle\langle 0|_E^{\otimes n}\otimes|w+nF_E\rangle\langle w+nF_E|_{W_0}\ , \tag{G.59}$$

recalling that $F_E = -\beta^{-1}\ln\big(\mathrm{tr}\big(e^{-\beta H_E}\big)\big)$. Similarly, there exists an energy-conserving unitary $K^{(2)}_{JW_0B_2}$ with an information battery J and a bath B_2 such that (in the limit of a very large bath),

$$\begin{aligned}\mathrm{tr}_{B_2}&\Big[K^{(2)}_{JW_0B_2}\big(\gamma_J\otimes|w+nF_E\rangle\langle w+nF_E|_{W_0}\otimes\gamma_{B_2}\big)K^{(2)\,\dagger}_{JW_0B_2}\Big]\\ &= (\mathbb{1}_m/m)_J\otimes\big|-\beta^{-1}\ln(d_J)\big\rangle\big\langle-\beta^{-1}\ln(d_J)\big|_{W_0}\ ,\end{aligned} \tag{G.60}$$

where $w + nF_E - w_J = -\beta^{-1}\ln(d_J)$, with $w_J = \beta^{-1}(\ln(d_J) - \ln(m))$ expressing the cost of this transformation. Also, there exists an energy-conserving unitary $K^{(3)}_{JW_0B_3}$ with an information battery J and a bath B_3 such that (in the limit of a very large bath),

$$\mathrm{tr}_{B_3}\Big[K^{(3)}_{JW_0B_3}\big(|0\rangle\langle 0|_J\otimes|-\beta^{-1}\ln(d_J)\rangle\langle-\beta^{-1}\ln(d_J)|_{W_0}\otimes\gamma_{B_3}\big)K^{(3)\,\dagger}_{JW_0B_3}\Big] = \gamma_J\otimes|0\rangle\langle 0|_{W_0}\ . \tag{G.61}$$

So now, we may construct the final, eventual, overall, total, ultimate unitary U_{X^nWB} of the claim. The bath B consists of the systems $JA^nB_1B_2B_3$, initialized as usual in their overall thermal state. Then, we set

$$\tilde{U}_{X^nWB} = |0\rangle_W\,\langle 0|_{W_0}\,K^{(3)}_{JW_0B_3}\,\tilde{W}_{E^nX^nA^nJ}\,K^{(2)}_{JW_0B_2}\,K^{(1)}_{E^nW_0B_1}|w\rangle_{W_0}\,\langle w|_W\ , \tag{G.62}$$

which is a unitary if the input state except for the input state that is fixed to $|w\rangle_W$ and the output state that is always $|0\rangle_W$. In fact $\tilde{U}_{X^nWB}$ can be completed to a full energy-conserving unitary by defining $U_{X^nWB} = \tilde{U}_{X^nWB} + \tilde{U}^\dagger_{X^nWB}$ (this works because of the two states $|0\rangle_W, |w\rangle_W$ that are orthogonal). It is easy to see that this unitary induces the same mapping $\Phi_{X^n\to X^n}$ with the appropriate input states, and hence fulfills all the claimed properties. ■

Appendix H: Optimal implementation of any i.i.d. channel with thermal operations on an i.i.d. input state

Here, we show that for any channel $\mathcal{E}_{X\to X'}$, mapping a system X with Hamiltonian H_X to a system X' with Hamiltonian $H_{X'}$, and for any fixed input state σ_X, then there exists an optimal implementation with thermal operations which uses a small amount of coherence, and an amount of work per copy which is asymptotically equal to

$$W_{\mathrm{T.O.}} = \beta^{-1}D(\mathcal{E}(\sigma)\,\|\,e^{-\beta H_{X'}}) - \beta^{-1}D(\sigma\,\|\,e^{-\beta H_X})\ . \tag{H.1}$$

The coherence is counted using a half-infinite energy latter with spacing x, i.e., with Hamiltonian $H_C = \sum_{k=0}^{d_C} kx|k\rangle\langle k|_C$, as considered in Ref. [34]. The state is initialized in the state $|\eta^L\rangle := L^{-1/2}\sum_{k=0}^{L-1}|\ell_0+k\rangle$, where ℓ_0 is a base energy offset. Such a system may be consumed entirely by the process, i.e., at the end of the process we may return it in any state we like. We assume that, in some reasonable setting, such a state on such a system can be prepared using an amount of work of the same order as the dimension of the system d_C. In the following, d_C will be sublinear in the number of copies n, so asymptotically for $n\to\infty$, the coherence source will require negligible resources to create when counted per copy. Actually, we will use two such sources C_1, C_2: We think of each of these as single-use disposable systems, and we need such systems at two stages in our protocol.

Proposition 22. *Let X, X' be systems, let $H_X, H_{X'}$ be the corresponding Hamiltonians, and let $\beta \geqslant 0$. Let $\mathcal{E}_{X\to X'}$ be a completely positive, trace-preserving map, and let σ_X be any input state. Let $R \simeq X$ and let $|\sigma\rangle_{XR} = \sigma_X^{1/2}\,|\Phi\rangle_{X:R}$ be a purification of σ_X. Then for any $0<\theta<1/3$, for any n and for any $\delta>0$ there exists a thermal operation which acts on $X\to X'$ and an information battery W whose process matrix is ϵ-close to $\mathcal{E}_{X\to X'}^{\otimes n}(\sigma_{XR}^{\otimes n})$ and which uses two coherence sources C_1, C_2 and an amount of work $W_{\mathrm{T.O.}}^n$, with*

$$\frac{1}{n}W_{\mathrm{T.O.}}^n \leqslant F(\rho_{X'}, H_{X'}) - F(\sigma_X, H_X) + O(n^{-1/2}) + O(n^{-1}\log(1/\theta)) \;; \tag{H.2a}$$

$$\epsilon \leqslant 3\theta + O(\exp(-cn\delta^2)) \;; \tag{H.2b}$$

$$d_{C_i}/n \leqslant O\big(\delta/\theta^2\big) \;, \tag{H.2c}$$

with $F(\sigma, H) = D(\sigma\,\|\,e^{-\beta H})$ and for some $c>0$ depending only on $H_X, H_{X'}$.

(Proof on page 47.)

The following corollary is obtained straightforwardly from the above proposition by choosing $\delta = n^{-1/2+\xi}$, $\theta = n^{-\xi/2}$ for any choice of $0<\xi<1/4$.

Corollary 23. *Any i.i.d. channel $\mathcal{E}_{X\to X'}^{\otimes n}$ between n copies of systems X, X' with Hamiltonians $H_X, H_{X'}$ can be implemented on a fixed i.i.d. input state $\sigma_X^{\otimes n}$ using thermal operations at a work cost rate per copy which is asymptotically equal to*

$$W_{\text{T.O., asympt.}} = F(\mathcal{E}(\sigma_X), H_{X'}) - F(\sigma_X, H_X) \;, \tag{H.3}$$

and using a vanishing amount of coherence per copy.

We need a technical lemma that tells us that whenever we have a Hamiltonian whose eigenvalues are all close, we may replace that Hamiltonian by an exactly flat Hamiltonian at a cost of investing an amount of coherence of the order of the energy spread we would like to flatten out.

Lemma 24 (Flattening Hamiltonians using [34])**.** *Let A be a system with a Hamiltonian H_A with m eigenvalues all lying in a range $[h_-, h_+]$. Let $B\simeq A$ be a system with a completely degenerate Hamiltonian $H_B = [(h_-+h_+)/2]\,\mathbb{1}_B$. We assume that the spacings of the eigenvalues of H_A as well as the value $(h_-+h_+)/2$ are multiple of some unit x. Fix $\theta>0$. Consider a coherence source C of dimension d_C, with Hamiltonian $H_C = \sum_{k=0}^{d_C} kx|k\rangle\langle k|_C$. Assume that the coherence source starts in the state $|\eta\rangle_C = L^{-1/2}\sum_{k=0}^{L-1}|\ell_0+k\rangle$ such that $\ell_0 \geqslant (h_-+h_+)/x$ and $L\geqslant \theta^{-2}(h_+-h_-)/x$ and $d_C \geqslant L+\ell_0+(h_-+h_+)/x$. Then there exists a partial isometry $U_{AC\to BC}$ which commutes exactly with the total Hamiltonians*

$(U_{AC\to BC}(H_A+H_C) = (H_B+H_C)U_{AC\to BC})$, such that for any ρ_{AR} on any reference system R, we have that $\rho_{AR}\otimes|\eta\rangle\langle\eta|_C$ is in the support of $U_{AC\to BC}\otimes\mathbb{1}_R$ and

$$P\Big(\mathrm{tr}_C\big(U\,(\rho_{AR}\otimes\eta)U^\dagger\big),\rho_{BR}\Big)\leqslant\theta\ , \tag{H.4}$$

where $\rho_{BR}=\mathrm{id}_{A\to B}(\rho_{AR})$ denotes the same state as the initial state, but on the systems BR.

The reverse operation $B\to A$ may also be carried out with the consumption of a similar coherence source, at the same accuracy.

Proof of Lemma 24. First, we can reduce the problem to a system dimension m: Embed the system into a bipartite system with a ancilla with trivial Hamiltonian storing the degeneracy index, and a second system with nondegenerate Hamiltonian storing the actual energy. Then the problem reduces to change the Hamiltonian of the second system. So we may assume without loss of generality that the Hamiltonian H_A is nondegenerate with $m=d_A$.

Consider the protocol of Åberg [34]. We would like to apply a result in the spirit of Åberg's Supplementary Proposition 2, but we need a tighter bound. Denoting the energy levels of A by $H_A=\sum xz_j|j\rangle\langle j|$ for integers z_j, we apply the global energy-conserving unitary on A and C given by

$$U_{AC\to BC}=\sum_j|j\rangle_B\langle j|_A\otimes\Delta^{z'-z_j}\ , \tag{H.5}$$

where $z'=(h_-+h_+)/(2x)$ represents the fixed energy of the output. Then, starting in the state $|\psi\rangle_{AA'}=\sum\psi_{jj'}|j\,j'\rangle_{AA'}$ using a reference system $A'\simeq A$ and the initial state $|\eta\rangle_C$ on C, we have

$$U_{AC\to BC}\,(|\psi\rangle_{AA'}|\eta\rangle_C)=\sum_{jj'}\psi_{jj'}|j\,j'\rangle_{BA'}\otimes(\Delta^{z'-z_j}|\eta\rangle_C)\ . \tag{H.6}$$

We may calculate the overlap with the initial state,

$$\mathrm{Re}\{\langle\psi|_{BA'}\langle\eta|_C\,U_{AC\to BC}\,|\psi\rangle_{AA'}|\eta\rangle_C\}=\sum_{jj'}|\psi_{jj'}|^2\,\mathrm{Re}\Big\{\langle\eta|_C\Delta^{z'-z_j}|\eta\rangle_C\Big\}\geqslant1-\frac{h_+-h_-}{xL}\ , \tag{H.7}$$

where we used the fact that $\mathrm{tr}(\Delta^{-a}\,\eta_C)=\max(0,1-a/L)$ and $|z'-z_j|\leqslant(h_+-h_-)/x$. Hence, using the fact that the partial trace can only increase the fidelity,

$$F^2\big(\Phi(|\psi\rangle\langle\psi|_{AA'}),|\psi\rangle\langle\psi|_{BA'}\big)\geqslant1-\frac{h_+-h_-}{xL}\ , \tag{H.8}$$

where $\Phi(\cdot)=\mathrm{tr}_C\big\{U_{AC\to BC}\,[(\cdot)\otimes|\eta\rangle\langle\eta|_C]\,U^\dagger\big\}$. Note that $|\psi\rangle_{AA'}$ is arbitrary at this point. For any state ρ_{AR}, using the joint concavity of the fidelity function, further noting that we may consider without loss of generality reference systems of the form $A'\simeq A$, we have

$$F^2(\Phi_{A\to B}(\rho_{AR}),\rho_{BR})\geqslant\min_{|\psi\rangle_{AA'}}F^2(\Phi_{A\to B}(|\psi\rangle\langle\psi|_{AA'}),|\psi\rangle\langle\psi|_{BA'})\geqslant1-\frac{h_+-h_-}{xL}\ , \tag{H.9}$$

and thus, since $L \geqslant \theta^{-2}(h_+ - h_-)/x$,

$$P(\Phi_{A\to B}(\rho_{AR}), \rho_{BR}) \leqslant \sqrt{\frac{h_+ - h_-}{xL}} \leqslant \theta \ . \tag{H.10}$$

The same argument can be applied to the operation $B \to A$. ■

Proof of Proposition 22. Let $\{R^k_{X^n}\}$ be the POVM elements from Proposition 5 for the input energy (for the Hamiltonian H_X) over the n systems, and let $\{L^\ell_{X'^n}\}$ be the corresponding output measurement.

We exhibit a protocol as a sequence of gentle measurements and thermal operations. For any $\delta > 0$, we measure the projector $R^{\approx_\delta \operatorname{tr}(H_X\sigma_X)}_{X^n}$ on the n inputs. This measurement fails with probability $\leqslant 2\exp(-\eta_1 n)$ for $\eta_1 = 2\delta^2/(\Delta H_X)^2$, where ΔH_X is the difference between the maximal and minimal eigenvalues of H_X.

Assume that the input and output Hamiltonians H_X and $H_{X'}$ have eigenvalues that are multiples of a spacing x (x may be very small). Now the state lies in a subspace of energies in the interval $[n(\operatorname{tr}(H_X\sigma_X) \pm \delta)]$. We invoke Lemma 24 to change this n-system Hamiltonian to one which is entirely flat, $H'_{X^n} := h\,\mathbb{1}_{X^n}$ with $h = n\operatorname{tr}(H_X\sigma_X)$. Given the target approximation parameter $\theta > 0$, the cost of this operation is the consumption of a coherence source C_1 of size $d_{C_1} = (\theta^{-2} + 2)\,(2n\delta/x) \leqslant O(n\delta/\theta^2)$, because $m = 2n\delta/x$.

Then we invoke the achievability result of Ref. [11], that one can implement any channel over trivial Hamiltonians using thermal operations, and an amount of work given approximately by the conditional entropy of the environment conditioned on the output. We use this step to implement the channel

$$S^{\approx_\delta \operatorname{tr}(H_{X'}\rho_{X'})}_{X'^n}\, \mathcal{E}^{\otimes n}_{X^n\to X'^n}(R^{\approx_\delta \operatorname{tr}(H_X\sigma_X)}_{X^n}\,(\cdot)\,R^{\approx_\delta \operatorname{tr}(H_X\sigma_X)}_{X^n})\, S^{\approx_\delta \operatorname{tr}(H_{X'}\rho_{X'})}_{X'^n} \tag{H.11}$$

up to an accuracy θ and investing an amount of work $nS(E\,|\,X')_\rho + O(\sqrt{n}) + \Delta(\theta)$ with $\Delta(\theta) = O(\log(1/\theta))$. We can now trivially shift the whole Hamiltonian $H'_{X^n} \to H'_{X'^n} := h'\mathbb{1}_{X'^n}$ with $h' = n\operatorname{tr}(H_{X'}\rho_{X'})$, investing an amount of work $h' - h$.

Finally, we have the state $S^{\approx_\delta \operatorname{tr}(H_{X'}\rho_{X'})}_{X'^n}\,\rho^{\otimes n}_{X'R}\,S^{\approx_\delta \operatorname{tr}(H_{X'}\rho_{X'})}_{X'^n}$ but the Hamiltonian is still the trivial $H'_{X'^n}$. Again we invoke Lemma 24 to change to the final Hamiltonian, up to accuracy θ, by consuming another coherence source C_2 of size $d_{C_2} \leqslant O(n\delta/\theta^2)$.

Because the final state is $S^{\approx_\delta \operatorname{tr}(H_{X'}\rho_{X'})}_{X'^n}\,\rho^{\otimes n}_{X'R}\,S^{\approx_\delta \operatorname{tr}(H_{X'}\rho_{X'})}_{X'^n}$ instead of $\rho^{\otimes n}_{X'R}$, we again pay a "gentle measurement penalty" of $O(e^{-n\eta_2/2})$ where $\eta_2 \sim \delta^2$ (cf. Lemma 27).

Finally, counting the total work, total failure probability and total use of coherence proves the claim, noting that $\operatorname{tr}(H_{X'}\rho_{X'}) - \operatorname{tr}(H_X\sigma_X) + S(E\,|\,X')_\rho = F(\rho_{X'}, H_{X'}) - F(\sigma_X, H_X)$. ■

Appendix I: Some technical lemmas

Lemma 25 (Diamond norm and trace distance)**.** *Let $\mathcal{K}_{X\to X'}, \mathcal{K}_{X\to X'}$ be quantum channels. Then*

$$\|\mathcal{K}_{X\to X'} - \mathcal{K}'_{X\to X'}\|_\diamond \leqslant d_X \, \|\mathcal{K}_{X\to X'}(\phi_{XR_X}) - \mathcal{K}_{X\to X'}(\phi_{XR_X})\|_1 \ , \tag{I.1}$$

where d_X is the dimension of X, where $R_X \simeq X$ is a reference system and where $|\phi\rangle_{X:R_X} = d_X^{-1/2} \sum_k |k\rangle_X |k\rangle_{R_X}$ is a maximally entangled state.

Proof of Lemma 25. We provide a simple proof for completeness. Let $|\sigma\rangle_{XR_X}$ be optimal for the diamond norm (the optimum can always be reached by a pure state), i.e.,

$$\|\mathcal{K}_{X\to X'} - \mathcal{K}'_{X\to X'}\|_\diamond = \|\mathcal{K}_{X\to X'}(\sigma_{XR_X}) - \mathcal{K}_{X\to X'}(\sigma_{XR_X})\|_1 \ . \tag{I.2}$$

We write $|\sigma\rangle_{X:R_X} = \sigma_{R_X}^{1/2} |\Phi\rangle_{X:R_X}$. By the properties of the one-norm (A.1), there exists a Hermitian Z_{XR_X} with $\|Z\|_\infty \leqslant 1$ such that

$$\begin{aligned}(\mathrm{I.2}) &= \mathrm{tr}\Big[Z_{XR_X}\, \sigma_{R_X}^{1/2}\, (\mathcal{K}_{X\to X'} - \mathcal{K}'_{X\to X'})(\Phi_{XR_X})\, \sigma_{R_X}^{1/2}\Big] \\ &= \mathrm{tr}\big[Z'_{XR_X}\, (\mathcal{K}_{X\to X'} - \mathcal{K}'_{X\to X'})(\Phi_{XR_X})\big] \ ,\end{aligned} \tag{I.3}$$

where we have defined $Z'_{XR_X} = \sigma_{R_X}^{1/2}\, Z_{XR_X}\, \sigma_{R_X}^{1/2}$. Noting that $\|Z'_{XR_X}\|_\infty \leqslant 1$, we have

$$(\mathrm{I.3}) \leqslant \|(\mathcal{K}_{X\to X'} - \mathcal{K}'_{X\to X'})(\Phi_{XR_X})\|_1 = d_X \, \|\mathcal{K}_{X\to X'}(\phi_{XR_X}) - \mathcal{K}_{X\to X'}(\phi_{XR_X})\|_1 \ , \tag{I.4}$$

as claimed. ∎

Lemma 26 (Pinching-like operator inequality)**.** *Let $\{E^i\}_{i=1}^M$ be a collection of M operators, and let $T \geqslant 0$. Then*

$$\Big(\sum E^i\Big)\, T \,\Big(\sum E^{j\,\dagger}\Big) \leqslant M \cdot \sum E^i\, T\, E^{i\,\dagger} \ . \tag{I.5}$$

Proof of Lemma 26. Call our system S and consider an additional register C of dimension $|C| = M$, and let $|\chi\rangle_C = M^{-1/2} \sum_{k=1}^M |k\rangle_C$. Then

$$\begin{aligned}\Big(\sum E_S^i\Big)\, T_S \,\Big(\sum E_S^{j\,\dagger}\Big) &= \mathrm{tr}_C\Big[\Big(\sum E_S^i \otimes |i\rangle_C\Big)\, T_S\, \Big(\sum E_S^{j\,\dagger} \otimes \langle j|_C\Big)\, (\mathbb{1}_S \otimes (M\, |\chi\rangle\langle\chi|_C))\Big] \\ &\leqslant M\, \mathrm{tr}_C\Big[\Big(\sum E_S^i \otimes |i\rangle_C\Big)\, T_S\, \Big(\sum E_S^{j\,\dagger} \otimes \langle j|_C\Big)\, (\mathbb{1}_S \otimes \mathbb{1}_C)\Big] \\ &= M \sum E_S^i\, T_S\, E_S^{i\,\dagger} \ ,\end{aligned} \tag{I.6}$$

using $|\chi\rangle\langle\chi|_C \leqslant \mathbb{1}_C$. ∎

Lemma 27 (Gentle measurement)**.** *Let ρ be any subnormalized quantum state and let $0 \leqslant Q \leqslant \mathbb{1}$. Suppose that $\operatorname{tr}(Q\rho) \geqslant 1-\delta$. Then*

$$P(\rho, Q^{1/2}\rho Q^{1/2}) \leqslant \sqrt{2\delta}\ . \tag{I.7}$$

(Note that this lemma is easier to prove than for instance [66, Lemma 7], thanks to the fact that we use a definition of the purified distance that differs slightly from that of Refs. [47, 56, 57, 66] if both states are subnormalized.)

Proof of Lemma 27. We have

$$F(\rho, Q^{1/2}\rho Q^{1/2}) = \operatorname{tr}\sqrt{\rho^{1/2}(Q^{1/2}\rho Q^{1/2})\rho^{1/2}} = \operatorname{tr}(Q^{1/2}\rho) \geqslant \operatorname{tr}(Q\rho) \geqslant 1-\delta\ . \tag{I.8}$$

Then $P(\rho, Q^{1/2}\rho Q^{1/2}) \leqslant \sqrt{1-(1-\delta)^2} \leqslant \sqrt{2\delta}$. ∎

Proposition 28. *Let W_X be an operator on a system X, such that $W^\dagger W \leqslant \mathbb{1}$. Then there exists a unitary operator U_{XQ} acting on X and a qubit Q such that, for any $|\psi\rangle_X$,*

$$\langle 0|_Q\, U_{XQ}\, (|\psi\rangle_X \otimes |0\rangle_Q) = W_X |\psi\rangle_X\ . \tag{I.9}$$

I.e., any operator W can be dilated to a unitary, with a post-selection on the output.

Proof of Proposition 28. Setting $V_{X\to XQ} = W \otimes |0\rangle_Q + \sqrt{\mathbb{1} - W^\dagger W} \otimes |1\rangle_Q$, we see that $V^\dagger V = W^\dagger W + \mathbb{1} - W^\dagger W = \mathbb{1}_X$, and hence $V_{X\to XQ}$ is an isometry. We can complete this isometry to a unitary U_{XQ} that acts as V on the support of $\mathbb{1}_X \otimes |0\rangle\langle 0|_Q$ and that maps the the support of $\mathbb{1}_X \otimes |1\rangle\langle 1|_Q$ onto the complementary space to the image of V. It then follows that for any $|\psi\rangle_X$, we have $U_{XQ}\,(|\psi\rangle_X \otimes |0\rangle_Q) = V_{X\to XQ}\,|\psi\rangle_X = (W_X|\psi\rangle_X) \otimes |0\rangle_Q + (\ldots) \otimes |1\rangle_Q$, and the claim follows. ∎

Proposition 29. *Let X be a quantum system with Hamiltonian H_X. Let W_X be an operator such that $W^\dagger W \leqslant \mathbb{1}$ and such that $[W_X, H_X] = 0$. Then there exists a unitary operator U_{XQ} acting on X and a qubit Q with $H_Q = 0$, that satisfies $[U_{XQ}, H_X] = 0$ and such that*

$$\langle 0|_Q\, U_{XQ}\, |0\rangle_Q = W_X\ , \tag{I.10}$$

I.e., any energy-preserving operator W can be dilated to an energy-preserving unitary on an ancilla with a post-selection on the output.

Proof of Proposition 29. First we calculate $[W^\dagger W, H_X] = W^\dagger[W, H_X] + [W^\dagger, H_X]\,W = 0 - [W, H_X]^\dagger W = 0$. This implies that $[\sqrt{\mathbb{1} - W^\dagger W}, H_X] = 0$, because $W^\dagger W$ and $\sqrt{\mathbb{1} - W^\dagger W}$ have the same eigenspaces. Define

$$V_{X\to XQ} = W \otimes |0\rangle_Q + \sqrt{\mathbb{1} - W^\dagger W} \otimes |1\rangle_Q\ . \tag{I.11}$$

The operator $V_{X\to XQ}$ is an isometry, because $V^\dagger V = W^\dagger W + \mathbb{1} - W^\dagger W = \mathbb{1}_X$. Furthermore, $V_{X\to XQ}\, H_X = (W_X H_X) \otimes |0\rangle + (\sqrt{\mathbb{1} - W^\dagger W}\, H_X) \otimes |1\rangle = (H_X W_X) \otimes |0\rangle + (H_X \sqrt{\mathbb{1} - W^\dagger W}) \otimes |1\rangle = H_X\, V_{X\to XQ}$ and thus $[V_{X\to XQ}, H_X] = 0$. Let $\{|j\rangle_X\}$ be an eigenbasis of H_X, and let $|\psi'_j\rangle_{XQ} = V_{X\to XQ}|j\rangle_X$, noting that both $|j\rangle_X$ and $|\psi'_j\rangle_{XQ}$ have the same energy. The two collections of vectors $\{|j\rangle_X \otimes |0\rangle_Q\}$ and $\{|\psi'_j\rangle_{XQ}\}$ can thus be completed into two bases $\{|\chi_i\rangle_{XQ}\}$ and $\{|\chi'_i\rangle_{XQ}\}$ of eigenvectors of $H_X + H_Q$ where the i-th element of both bases have the same energy. Define finally $U_{XQ} = \sum_i |\chi'_i\rangle\langle\chi_i|_{XQ}$, noting that by construction $U_{XQ}|0\rangle_Q = V_{X\to XQ}$ and $[U_{XQ}, H_X] = 0$. ■

Proposition 30. *Let X be a quantum system with Hamiltonian H_X, and let W_X be any operator satisfying $W^\dagger W \leqslant \mathbb{1}$, and such that $[W_X, H_X] = 0$. Then for any $\epsilon > 0$, there exists a unitary operator U_X satisfying $[U_X, H_X] = 0$ and such that for any states $|\psi\rangle_X, |\psi'\rangle_X$ satisfying $\mathrm{Re}\{\langle\psi'|W|\psi\rangle\} \geqslant 1-\epsilon$, we have $\mathrm{Re}\{\langle\psi'|U|\psi\rangle\} \geqslant 1 - 6\epsilon^{1/4}$.*

Proof of Proposition 30. Let $F = W^\dagger W \leqslant \mathbb{1}$ noting that $F^\dagger = F$. For some ν with $0 < \nu < 1$ to be determined later, let P be the projector onto the eigenspaces of F corresponding to eigenvalues greater or equal to ν. Define $V = WF^{-1/2}P$. The operator V is a partial isometry, meaning that its singular values are all equal to one or to zero, because $V^\dagger V = PF^{-1/2}W^\dagger WF^{-1/2}P = PF^{-1/2}FF^{-1/2}P = P$, since P lies within the support of F. Observe that $[F,H] = [W^\dagger W, H] = W^\dagger[W,H] - [W,H]^\dagger W = 0$ and hence $[F^{-1/2}, H] = 0$ and $[P,H] = 0$. Hence, $[V,H] = [WF^{-1/2}P, H] = 0$. So we may complete the partial isometry V into a full unitary U that also commutes with H by acting as the identity on the remaining elements of the eigenbasis of H in which V is diagonal. We may thus write $U = V + X = WF^{-1/2}P + X$ for some operator X satisfying $XP = 0$ and $[X,H] = 0$.

Now let $|\psi\rangle$, $|\psi'\rangle$ such that $\mathrm{Re}\{\langle\psi'|W|\psi\rangle\} \geqslant 1-\epsilon$, and write

$$\begin{aligned}\mathrm{Re}\{\langle\psi'|U|\psi\rangle\} &= \mathrm{Re}\{\langle\psi'|(U-W)|\psi\rangle\} + \mathrm{Re}\{\langle\psi'|W|\psi\rangle\}\\ &\geqslant \mathrm{Re}\{\langle\psi'|(U-W)|\psi\rangle\} + 1 - \epsilon\ .\end{aligned} \tag{I.12}$$

We have $\langle\psi|P|\psi\rangle \geqslant \langle\psi|PFP|\psi\rangle = \langle\psi|F|\psi\rangle - \langle\psi|(\mathbb{1}-P)F|\psi\rangle \geqslant \langle\psi|F|\psi\rangle - \nu$, recalling that $\mathbb{1} - P$ projects onto the eigenspaces of F whose eigenvalues are less than ν. Then, $\langle\psi|F|\psi\rangle = \langle\psi|W^\dagger W|\psi\rangle \geqslant \langle\psi|W^\dagger|\psi'\rangle\langle\psi'|W|\psi\rangle = |\langle\psi'|W|\psi\rangle|^2 \geqslant (1-\epsilon)^2 \geqslant 1-2\epsilon$, and hence $\left\|(\mathbb{1}-P)|\psi\rangle\right\|^2 = \langle\psi|(\mathbb{1}-P)|\psi\rangle \leqslant 1 - (\langle\psi|F|\psi\rangle - \nu) \leqslant 1 - (1-2\epsilon) + \nu = 2\epsilon + \nu$. Hence

$$\begin{aligned}\mathrm{Re}\big\{\langle\psi'|(U-W)|\psi\rangle\big\} &= \mathrm{Re}\big\{\langle\psi'|(U-W)P|\psi\rangle\big\} + \mathrm{Re}\big\{\langle\psi'|(U-W)(\mathbb{1}-P)|\psi\rangle\big\}\\ &\geqslant \mathrm{Re}\big\{\langle\psi'|(U-W)P|\psi\rangle\big\} - 2\sqrt{2\epsilon+\nu}\ ,\end{aligned} \tag{I.13}$$

since by Cauchy-Schwarz $\left|\langle\psi'|(U-W)(\mathbb{1}-P)|\psi\rangle\right| \leqslant \left\|(U-W)^\dagger|\psi'\rangle\right\| \left\|(\mathbb{1}-P)|\psi\rangle\right\|$, where $\|(U-W)^\dagger|\psi'\rangle\| \leqslant 2$. In order to continue, let $|\chi\rangle = W^\dagger|\psi'\rangle - |\psi\rangle$, and calculate $\langle\chi|\chi\rangle = \langle\psi'|WW^\dagger|\psi'\rangle + \langle\psi|\psi\rangle - 2\,\mathrm{Re}\{\langle\psi'|W|\psi\rangle\} \leqslant 2 - 2(1-\epsilon) = 2\epsilon$, and hence we deduce that $\left\||\chi\rangle\right\| = \left\|W^\dagger|\psi'\rangle - |\psi\rangle\right\| \leqslant \sqrt{2\epsilon}$. Then, with $\langle\psi'|W = \langle\psi| + \langle\chi|$ we have

$$\begin{aligned}\mathrm{Re}\big\{\langle\psi'|(U-W)P|\psi\rangle\big\} &= \mathrm{Re}\big\{\langle\psi'|W(F^{-1/2}-\mathbb{1})P|\psi\rangle\big\}\\ &= \mathrm{Re}\big\{\langle\chi|(F^{-1/2}-\mathbb{1})P|\psi\rangle\big\} + \mathrm{Re}\big\{\langle\psi|(F^{-1/2}-\mathbb{1})P|\psi\rangle\big\}\\ &\geqslant \mathrm{Re}\big\{\langle\chi|(F^{-1/2}-\mathbb{1})P|\psi\rangle\big\}\ ,\end{aligned} \tag{I.14}$$

since $\langle\psi|(F^{-1/2}-\mathbb{1})P|\psi\rangle = \langle\psi|P(F^{-1/2}-\mathbb{1})P|\psi\rangle \geqslant 0$ as P commutes with $F^{-1/2}$ and since $F = W^\dagger W \leqslant \mathbb{1}$ implies that $F^{-1/2} \geqslant \mathbb{1}$. To bound the remaining term we first write $\left|\langle\chi|(F^{-1/2}-\mathbb{1})P|\psi\rangle\right| \leqslant \left\||\chi\rangle\right\|\left\|(F^{-1/2}-\mathbb{1})P|\psi\rangle\right\| \leqslant \sqrt{2\epsilon/\nu}$; the last inequality follows since P projects onto the eigenspaces of F with eigenvalues larger than or equal to ν, thus $F^{-1/2}P \leqslant \nu^{-1/2}P$ and hence $\left\|(F^{-1/2} - \mathbb{1})P|\psi\rangle\right\| \leqslant \nu^{-1/2} - 1 \leqslant \nu^{-1/2}$. Hence,

$$\text{(I.14)} \geqslant -\sqrt{\frac{2\epsilon}{\nu}} \tag{I.15}$$

Following the inequalities from (I.12), invoking (I.13) and with the above, we finally obtain

$$\mathrm{Re}\{\langle\psi'|U|\psi\rangle\} \geqslant 1 - \epsilon - 2\sqrt{2\epsilon + \nu} - \sqrt{\frac{2\epsilon}{\nu}}\ . \tag{I.16}$$

Choosing $\nu = 2\epsilon^{1/2}$, we obtain, using $\epsilon \leqslant \sqrt{\epsilon}$,

$$1 - \mathrm{Re}\{\langle\psi'|U|\psi\rangle\} \leqslant \epsilon + 2\sqrt{2\epsilon + 2\epsilon^{1/2}} + \sqrt{\frac{2\epsilon}{2\epsilon^{1/2}}} \leqslant (1+4+1)\,\epsilon^{1/4} = 6\,\epsilon^{1/4}\ , \tag{I.17}$$

as claimed. ■

Proposition 31 (Controlled-unitary using a POVM)**.** *Let $\{Q^j\}$ be a set of positive semidefinite operators on a system X satisfying $\sum Q^j \leqslant \mathbb{1}$, and let $\{U^j\}$ be a collection of unitaries on a system Y. Let*

$$W_{XY} = \sum_j Q_X^j \otimes U_Y^j\ . \tag{I.18}$$

Then $W^\dagger W \leqslant \mathbb{1}$.

Proof of Proposition 31. Using an additional register K, define

$$V_{X\to XK} = \sum \sqrt{Q^j} \otimes |j\rangle_K\ . \tag{I.19}$$

Then $V^\dagger V = \sum Q^j \leqslant \mathbb{1}$. Clearly, $VV^\dagger \leqslant \mathbb{1}_{XK}$ because $VV^\dagger$ and $V^\dagger V$ have the same nonzero eigenvalues. Now observe that

$$W = V^\dagger\Big(\sum \mathbb{1}_X \otimes U_Y^j \otimes |j\rangle\langle j|_K\Big)\,V\ , \tag{I.20}$$

where the middle term is clearly unitary, and hence manifestly we have $W^\dagger W \leqslant \mathbb{1}$. ■

[1] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, "Resource theory of quantum states out of thermal equilibrium," Physical Review Letters **111**, 250404 (2013), arXiv:1111.3882.

[2] M. Horodecki and J. Oppenheim, "Fundamental limitations for quantum and nanoscale thermodynamics," Nature Communications **4**, 2059 (2013), arXiv:1111.3834.

[3] F. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, "The second laws of quantum thermodynamics," Proceedings of the National Academy of Sciences **112**, 3275–3279 (2015), arXiv:1305.5278.

[4] G. Gour, D. Jennings, F. Buscemi, R. Duan, and I. Marvian, "Quantum majorization and a complete set of entropic conditions for quantum thermodynamics," (2017), arXiv:1708.04302.

[5] C. T. Chubb, M. Tomamichel, and K. Korzekwa, "Beyond the thermodynamic limit: finite-size corrections to state interconversion rates," (2017), arXiv:1711.01193.

[6] P. Faist and R. Renner, "Fundamental work cost of quantum processes," Physical Review X **8**, 021011 (2018), arXiv:1709.00506.

[7] O. C. O. Dahlsten, R. Renner, E. Rieper, and V. Vedral, "Inadequacy of von Neumann entropy for characterizing extractable work," New Journal of Physics **13**, 53015 (2011), arXiv:0908.0424.

[8] J. Åberg, "Truly work-like work extraction via a single-shot analysis," Nature Communications **4**, 1925 (2013), arXiv:1110.6121.

[9] L. del Rio, J. Åberg, R. Renner, O. Dahlsten, and V. Vedral, "The thermodynamic meaning of negative entropy," Nature **474**, 61–63 (2011), arXiv:1009.1630.

[10] M. Navascués and L. P. García-Pintos, "Nonthermal quantum channels as a thermodynamical resource," Physical Review Letters **115**, 010405 (2015), arXiv:1501.02597.

[11] P. Faist, F. Dupuis, J. Oppenheim, and R. Renner, "The minimal work cost of information processing," Nature Communications **6**, 7669 (2015), arXiv:1211.1037.

[12] A. W. Harrow, *Applications of coherent classical communication and the Schur transform to quantum information theory*, Ph.D. thesis, Massachusetts Institute of Technology (2005), arXiv:quant-ph/0512255.

[13] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, "The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels," IEEE Transactions on Information Theory **60**, 2926–2959 (2014), arXiv:0912.5537.

[14] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, "Sample-optimal tomography of quantum states," IEEE Transactions on Information Theory **63**, 5628–5641 (2017), arXiv:1508.01797.

[15] M. Christandl, R. König, and R. Renner, "Postselection technique for quantum channels with applications to quantum cryptography," Physical Review Letters **102**, 20504 (2009), arXiv:0809.3019.

[16] R. Jozsa, M. Horodecki, P. Horodecki, and R. Horodecki, "Universal quantum information compression," Physical Review Letters **81**, 1714–1717 (1998), arXiv:quant-ph/9805017.

[17] M. Hayashi and K. Matsumoto, "Quantum universal variable-length source coding," Physical Review A **66**, 022311 (2002), arXiv:quant-ph/0209124.

[18] R. Jozsa and S. Presnell, "Universal quantum information compression and degrees of prior knowledge," Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **459**, 3061–3077 (2003), arXiv:quant-ph/0210196.

[19] C. H. Bennett, A. W. Harrow, and S. Lloyd, "Universal quantum data compression via nondestructive tomography," Physical Review A **73**, 032336 (2006), arXiv:quant-ph/0403078.

[20] I. Bjelakovic and R. Siegmund-Schultze, "Quantum Stein's lemma revisited, inequalities for quantum entropies, and a concavity theorem of Lieb," (2003), arXiv:quant-ph/0307170.

[21] J. Nötzel, "A solution to two party typicality using representation theory of the symmetric group," (2012), arXiv:1209.5094.

[22] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, 2013) arXiv:1106.1445.

[23] M. Berta, M. Lemm, and M. M. Wilde, "Monotonicity of quantum relative entropy and recoverability," Quantum Information and Computation **15**, 1333–1354 (2015), arXiv:1412.4067.

[24] P. Sen, "A one-shot quantum joint typicality lemma," (2018), arXiv:1806.07278.

[25] A. Anshu, V. K. Devabathini, and R. Jain, "Quantum communication using coherent rejection

sampling," Physical Review Letters **119**, 120506 (2017), arXiv:1410.3031.

[26] A. Anshu, R. Jain, and N. A. Warsi, "One shot entanglement assisted classical and quantum communication over noisy quantum channels: A hypothesis testing and convex split approach," (2017), arXiv:1702.01940.

[27] A. Anshu, R. Jain, and N. A. Warsi, "A one-shot achievability result for quantum state redistribution," IEEE Transactions on Information Theory **64**, 1425–1435 (2018), arXiv:1702.02396.

[28] A. Anshu, R. Jain, and N. A. Warsi, "A generalized quantum Slepian–Wolf," IEEE Transactions on Information Theory **64**, 1436–1453 (2018), arXiv:1703.09961.

[29] A. Anshu, R. Jain, and N. A. Warsi, "One-shot measurement compression with quantum side information using shared randomness," (2017), arXiv:1703.02342.

[30] J. Åberg, "Fully quantum fluctuation theorems," Physical Review X **8**, 011019 (2018), arXiv:1601.01302.

[31] Á. M. Alhambra, L. Masanes, J. Oppenheim, and C. Perry, "Fluctuating work: From quantum thermodynamical identities to a second law equality," Physical Review X **6**, 041017 (2016), arXiv:1601.05799.

[32] M. Berta, M. Christandl, and R. Renner, "The quantum reverse Shannon theorem based on one-shot information theory," Communications in Mathematical Physics **306**, 579–615 (2011), arXiv:0912.3805.

[33] I. Marvian and R. W. Spekkens, "Asymmetry properties of pure quantum states," Physical Review A **90**, 014102 (2014), arXiv:1105.1816.

[34] J. Åberg, "Catalytic coherence," Physical Review Letters **113**, 150402 (2014), arXiv:1304.1060.

[35] M. Lostaglio, K. Korzekwa, D. Jennings, and T. Rudolph, "Quantum coherence, time-translation symmetry, and thermodynamics," Physical Review X **5**, 021001 (2015), arXiv:1410.4572.

[36] C. Cîrstoiu and D. Jennings, "Irreversibility and quantum information flow under global & local gauge symmetries," (2017), arXiv:1707.09826.

[37] I. Marvian, "Coherence distillation machines are impossible in quantum thermodynamics," (2018), arXiv:1805.01989.

[38] C. Majenz, M. Berta, F. Dupuis, R. Renner, and M. Christandl, "Catalytic decoupling of quantum information," Physical Review Letters **118**, 080503 (2017), arXiv:1605.00514.

[39] L. Wang and R. Renner, "One-shot classical-quantum capacity and hypothesis testing," Physical Review Letters **108**, 200501 (2012), arXiv:1007.5456.

[40] M. Tomamichel and M. Hayashi, "A hierarchy of information quantities for finite block length analysis of quantum tasks," IEEE Transactions on Information Theory **59**, 7693–7710 (2013), arXiv:1208.1478.

[41] W. Matthews and S. Wehner, "Finite blocklength converse bounds for quantum channels," IEEE Transactions on Information Theory **60**, 7317–7329 (2014), arXiv:1210.4722.

[42] F. Buscemi and N. Datta, "The quantum capacity of channels with arbitrarily correlated noise," IEEE Transactions on Information Theory **56**, 1447–1460 (2010), arXiv:0902.0158.

[43] F. G. S. L. Brandão and N. Datta, "One-shot rates for entanglement manipulation under non-entangling maps," IEEE Transactions on Information Theory **57**, 1754–1760 (2011), arXiv:0905.2673.

[44] F. Dupuis, L. Kraemer, P. Faist, J. M. Renes, and R. Renner, "Generalized entropies," in *XVIIth International Congress on Mathematical Physics* (2013) pp. 134–153, arXiv:1211.3141.

[45] R. Renner, *Security of Quantum Key Distribution*, Ph.D. thesis, ETH Zürich (2005), arXiv:quant-ph/0512258.

[46] N. Datta, "Min- and max-relative entropies and a new entanglement monotone," IEEE Transactions on Information Theory **55**, 2816–2826 (2009), arXiv:0803.2770.

[47] M. Tomamichel, *A Framework for Non-Asymptotic Quantum Information Theory*, Ph.D. thesis, ETH Zurich (2012), arXiv:1203.2142.

[48] P. Faist, J. Oppenheim, and R. Renner, "Gibbs-preserving maps outperform thermal operations in the quantum regime," New Journal of Physics **17**, 043003 (2015), arXiv:1406.3618.

[49] M. Berta, J. M. Renes, and M. M. Wilde, "Identifying the information gain of a quantum measurement," IEEE Transactions on Information Theory **60**, 7987–8006 (2014), arXiv:1301.1594.

[50] S. P. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, UK, 2004).

[51] A. Capel, A. Lucia, and D. Pérez-García, "Quantum conditional relative entropy and quasi-factorization of the relative entropy," (2018), arXiv:1804.09525.

[52] K. Fang, X. Wang, M. Tomamichel, and M. Berta, "Quantum channel simulation and the channel's smooth max-information," (2018), to appear.

[53] A. Anshu, M. Berta, R. Jain, and M. Tomamichel, "Partially smoothed information measures," (2018), to appear.

[54] N. Dutil, *Multiparty quantum protocols for assisted entanglement distillation*, Ph.D. thesis, McGill University, Montréal (2011), arXiv:1105.4657.

[55] L. Drescher and O. Fawzi, "On simultaneous min-entropy smoothing," in *2013 IEEE International Symposium on Information Theory* (IEEE, 2013) pp. 161–165, arXiv:1312.7642.

[56] M. Tomamichel, R. Colbeck, and R. Renner, "Duality between smooth min- and max-entropies," IEEE Transactions on Information Theory **56**, 4674–4681 (2010), arXiv:0907.5238.

[57] M. Tomamichel, *Quantum Information Processing with Finite Resources*, SpringerBriefs in Mathematical Physics, Vol. 5 (Springer International Publishing, Cham, 2016) arXiv:1504.00233.

[58] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).

[59] J. Watrous, "Semidefinite programs for completely bounded norms," Theory of Computing **5**, 217–238 (2009), arXiv:0901.4709.

[60] L. Szilard, "Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen," Zeitschrift für Physik **53**, 840–856 (1929).

[61] A. Pitchford, C. Granade, P. D. Nation, and R. J. Johansson, "QuTip 4.1.0," (2016).

[62] J.R. Johansson, P.D. Nation, and F. Nori, "QuTip 2: A Python framework for the dynamics of open quantum systems," Computer Physics Communications **184**, 1234–1240 (2013), arXiv:1211.6518.

[63] M. S. Andersen, J. Dahl, and L. Vandenberghe, "CVXOPT 1.1.9," (2016).

[64] K. M. R. Audenaert and M. Mosonyi, "Upper bounds on the error probabilities and asymptotic error exponents in quantum multiple state discrimination," Journal of Mathematical Physics **55**, 102201 (2014), arXiv:1401.7658.

[65] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," IEEE Transactions on Information Theory **49**, 1753–1768 (2003), arXiv:quant-ph/0206186.

[66] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner, "The uncertainty principle in the presence of quantum memory," Nature Physics **6**, 659–662 (2010), arXiv:0909.0950.