

RANDOMNESS, CATALYSIS  
AND PARTIAL KNOWLEDGE  
IN QUANTUM THERMODYNAMICS

*im Fachbereich Physik der Freien Universität Berlin eingereichte Dissertation  
zur Erlangung des Grades eines Doktors der Naturwissenschaften*

PAUL BOËS

*Berlin, 2020*

Erstgutachter: Prof. Dr. Jens Eisert  
Zweitgutachter: Prof. Felix von Oppen, PhD.  
Tag der Disputation: 15. Februar 2021

*Dedicated to  
Gebhard (-2018) and Marleen (2019-)*



## CONTENTS

---

1	SUMMARY	1
2	LIST OF PUBLICATIONS	3
3	INTRODUCTION	5
3.1	Quantum Thermodynamics	5
3.2	From the Schrödinger equation to the Second Law	7
3.3	Catalysis, Embezzlement, Coherences	13
4	CATALYTIC QUANTUM RANDOMNESS	19
4.1	Random unitary channels and thermal operations	19
4.2	Majorization	21
4.3	Randomness as a resource	22
5	CORRELATED CATALYSIS	45
5.1	Motivation and characterization	45
5.1.1	Reusing catalysts on sequences of systems	45
5.1.2	Characterization via von Neumann entropy	46
5.2	Catalysis for fluctuation theorems	59
5.2.1	Fluctuation theorems	59
5.2.2	Bypassing fluctuation theorems with correlated catalysis	60
6	PARTIAL INFORMATION AND THE CANONICAL ENSEMBLE	77
6.1	Thermal operations under partial information	77
6.2	Roads to the canonical ensemble	77
6.3	Statistical ensembles without typicality	79
7	CONCLUSIONS	105
7.1	Open questions	105
7.2	Acknowledgements	107
8	BIBLIOGRAPHY	109
9	BACK MATTER	119
9.1	Zusammenfassung	119
9.2	Anteile des Autors an zugrundeliegenden Arbeiten	120
9.3	Selbstständigkeitserklärung	121



## SUMMARY

---

Quantum thermodynamics is a blossoming research program that uses tools and ideas from quantum information theory to extend the laws of thermodynamics to the domain of systems to which the laws of quantum mechanics apply and the thermodynamic limit does not necessarily apply. The contributions of this program are both foundational, in providing an approach to constructively derive the laws of phenomenological thermodynamics from the postulates of unitary quantum mechanics in a rigorous and bottom-up fashion, but also practical, in providing the theory for an increasing number of experiments that attempt to build thermodynamic machines at scales that were previously inaccessible.

One key mathematical tool used in quantum thermodynamics is the framework of resource theories, specifically the resource theory of thermal operations, in which the thermodynamic interaction of a system with a heat bath and various additional components such as work batteries, clocks or catalysts are modeled. In this cumulative thesis, various extensions and modifications of this framework are studied in order to derive novel results and insights about the possible thermodynamic evolution of quantum systems. In particular, following a systematic exposition of the resource theory of thermal operations from first principles, I develop answers to the following questions: i) What is the smallest heat bath required to provide the necessary randomness for a system to undergo a given stochastic or thermodynamical evolution? ii) Does this size differ depending on whether the interaction between bath and system is quantum or classical? iii) How can quantum systems thermodynamically evolve in the presence of catalytic bystander systems and how can this be put to use? iv) What are the thermodynamic state transitions that an agent can operationally affect when she only has access to partial information about the underlying states of system and bath? v) How can we understand the emergence of the canonical ensemble in statistical thermodynamics from quantum mechanics?

These questions cover a wide ground, but it will become clear that they can in fact all be discussed using the same formal tools. As such, the answers to the above questions provided in this thesis are both interesting in their own right – showing for example that there exists a gap in how efficiently randomness can be exploited quantumly as compared to classically or that catalytic bystander systems can be used to extract finite amounts of work per particle from macroscopic systems with non-vanishing probability – as well as illustrate the power of quantum thermodynamics as a set of tools to connect quantum mechanics and the theory of thermodynamics more generally.





## LIST OF PUBLICATIONS

This cumulative thesis contains the following first-author publications:

- [1] Paul Boes, Henrik Wilming, Rodrigo Gallego, and Jens Eisert  
*Catalytic Quantum Randomness*,  
**Phys. Rev. X** **8**, 041016 (2018), DOI: 10.1103/PhysRevX.8.041016
- [2] Paul Boes, Jens Eisert, Rodrigo Gallego, Markus P. Müller, and Henrik Wilming  
*Von Neumann Entropy from Unitarity*,  
**Phys. Rev. Lett.** **122**, 210402 (2019), DOI: 10.1103/PhysRevLett.122.210402
- [3] Paul Boes, Rodrigo Gallego, Nelly Ng, Jens Eisert, and Henrik Wilming  
*By-passing fluctuation theorems with a catalyst*,  
**Quantum** **4**, 231 (2020) DOI: 10.22331/q-2020-02-20-231
- [4] Paul Boes, Henrik Wilming, Jens Eisert, and Rodrigo Gallego  
*Statistical ensembles without typicality*,  
**Nature Communications** **Vol. 9**, 1022 (2018), DOI: 10.1038/s41467-018-03230-y

Furthermore, during the completion of this doctoral research, the following other works were produced:

- [5] Alexander Streltsov, Swapan Rana, Paul Boes, and Jens Eisert  
*Structure of the resource theory of quantum coherence*,  
**Phys. Rev. Lett.** **119**, 140402 (2017), DOI: 10.1103/PhysRevLett.119.140402
- [6] Paul Boes, and Miguel Navascues  
*Composing decoherence functionals*  
**Phys. Rev. A** **95**, 022114 (2017), DOI: 10.1103/PhysRevA.95.022114
- [7] Carlo Sparaciari, Marcel Goihl, Paul Boes, Jens Eisert, and Nelly Ng  
*Bounding the resources for thermalizing many-body localized systems*  
**arXiv:1904.01314**
- [8] Paul Boes, Henrik Wilming, Nelly Ng and Jens Eisert  
*Second-order constraints for single-shot resource theories*  
**in preparation**
- [9] Henrik Wilming and Paul Boes  
*Separation of unital, exactly factorizable and random unitary channel via catalytic dilations*  
**in preparation**



## INTRODUCTION

---

### 3.1 QUANTUM THERMODYNAMICS

Phenomenological thermodynamics, that is, the theory of thermodynamics that was developed, amongst other people, by Sadi Carnot, William Thomson (a.k.a. Lord Kelvin) and Rudolf Clausius in the first half of the 19th century, is what Albert Einstein called a *principal* theory [10]. Principal theories, which Einstein distinguished from *constructive* theories, are theories whose laws are empirically well-confirmed statements — principles —, that constrain the possible states and evolutions of systems that fall under the domain of applicability of the theory. Phenomenological thermodynamics is a principal theory because its laws, in particular the First, Second and Third Law, have been formulated in reaction to experiments that the above researchers and others carried out and which the latter have found to hold in many different experimental settings, without being able to, and necessarily seeking to, produce a transparent explanation for *why* these laws would hold. Such explanations, Einstein holds, are not provided by principal theories but instead by constructive theories — such as Newtonian mechanics — whose laws are derived, bottom-up, from a set of axioms. Despite their lack of explanatory power, principal theories can be useful as guides in the development of scientific theories, especially when no constructive theory to understand a given set of phenomena exists, or when reducing a phenomenon to the underlying workings of a constructive theory is mathematically not tractable .

That the concepts of phenomenological thermodynamics have turned out to be useful guides for the development of theories in physics is beyond doubt, with the notions of entropy, temperature and the laws of phenomenological thermodynamics having played important roles in the development of the theories of black holes [11, 12], biological physics [13], many-body physics and statistical physics [14], information theory [15], computer science [16, 17, 18] and other fields. Nevertheless, it has from the start been a concern for researchers to provide constructive explanations for the laws of thermodynamics, by deriving them from a constructive theory of the dynamics of the microscopic constituents of thermodynamic systems. Prominent efforts, culminating in the development of statistical thermodynamics, here include those of James Maxwell, Ludwig Boltzmann and Josiah Gibbs (see [19] for an excellent historical review of these efforts). The intent of these reductional efforts was both foundational — for example to better understand how a time-asymmetric theory such as phenomenological thermodynamics could result from the underlying workings of time-symmetric Newtonian mechanics — but also to extend the laws of thermodynamics beyond their original domain of applicability.

Fast-forwarding 150 years, physics has moved a long way in better understanding the relationship between thermodynamics and the microscopic dynamics of matter. Mile-stones include fluctuation theorems [20, 21, 22, 23, 24], the development of the notion of typicality and concentration inequalities (e.g.[25, 26]), an improved understanding of the processes of equilibration and thermalization in physical systems [27] and their absence [28, 29]. Also, on the experimental side (see [30, 31, 32, 33] for examples), there have been great advances in the size of systems that can be controlled, with experimenters being able, for example, to deploy heat engines that using “heat baths” composed of clouds of  $10^5$  atoms and working

materials composed of  $10^4$  atoms [30]. Still, there remain open questions in the reduction of thermodynamics to quantum mechanics. Moreover, the improved experimental access requires an extension of the laws of thermodynamics to scales at which quantum effects become relevant.

Quantum thermodynamics, as I mean it here<sup>1</sup>, is a relatively young research program, in which tools from quantum information theory are used to extend the laws of thermodynamics, and their applications, to quantum systems outside of the thermodynamic limit. These systems might be very small quantum systems, but they need not be. The laws of phenomenological thermodynamics are derived as special cases from those of quantum thermodynamics, but they also are valid in settings in which the latter are not well-defined, and hence expand the domain of applicability of thermodynamics.

In this thesis, I use the tools of quantum thermodynamics to study and answer the following questions:

1. What is the smallest size of a heat bath required to provide the necessary randomness for a system to undergo a given stochastic or thermodynamical evolution?
2. Does this size differ depending on whether the interaction between bath and system is quantum or classical?
3. How can quantum systems thermodynamically evolve in the presence of catalytic bystander systems?
4. What are the thermodynamic state transitions that an agent can operationally affect when she only has access to partial information about the underlying states of system and bath?
5. How can we understand the emergence of the canonical ensemble in statistical thermodynamics from quantum mechanics?

I do so as follows: In the remainder of the introduction, I will introduce the *resource theory of (catalytic) thermal operations* [38] as the formal starting point for the study of the above questions, and also clarify the connection of the theory of thermal operations to a theory of random processes. Chapter 4 addresses questions 1 and 2. In particular, I formalize how to measure the amount of randomness that is provided by (part of) a heat bath required to implement a random process and show that there is a gap in the amount of randomness required when the interaction with the bath that induces a random evolution of a system is quantum-mechanical or classical. Chapter 5 addresses question 3 by introducing the notion of correlated catalysis. Correlated catalysis provides a natural generalization of the more common notion of uncorrelated catalysis, in which a catalytic ancilla has to be returned both locally unchanged and uncorrelated from the system with which it interacted. In correlated catalysis this latter constraint is dropped. This turns out to provide setups that are of significant mathematical and physical interest. In particular, mathematically, correlated catalysis is shown to operationally single out the von Neumann entropy as a *single-shot* quantity, which challenges the folklore knowledge that the von Neumann entropy only is a special entropic quantity in the i. i. d. limit (Sec. 5.1). Moreover, physically, correlated catalysis can, in principle, be used

<sup>1</sup> Unsurprisingly, there are other research programs that also go by the name of quantum thermodynamics, but whose methodology and basic vocabulary differs considerably from that of quantum thermodynamics as I present it in this thesis (especially in the application of tools from quantum information theory). The works [34, 35, 36] provide a good introduction to quantum thermodynamics as I use the term here. For an example of “quantum thermodynamics” that differs from what I discuss in this thesis, see [37].

to bypass the constraints imposed by fluctuation theorems to extract work from macroscopic systems in thermal equilibrium with non-vanishing probability (Sec. 5.2). Finally, Chapter 6 addresses questions 4 and 5. In particular, tools are developed that provide a novel, operational motivation for the use of the canonical ensemble as a representation of a system in thermodynamic equilibrium. In particular, it is shown that the canonical ensemble is the only state that captures the state transitions that an experimenter with only partial knowledge of and partial control over their system can implement. As such, canonical ensembles are not only special from a mathematical point of view (as shown for instance by typicality and passivity approaches) but they also reflect the practical aspects of dealing with thermodynamic systems of which one holds only partial information.

### 3.2 FROM THE SCHRÖDINGER EQUATION TO THE SECOND LAW

As mentioned before, quantum thermodynamics is concerned with how the laws of phenomenological thermodynamics can be rigorously derived from the postulates of unitary quantum mechanics by using concepts from quantum information theory. As such, the starting point for the resource theory of catalytic thermal operations (CTO) are density operators on Hilbert spaces as representations of mixed quantum states, the tensor product structure for Hilbert spaces of joint systems and the Schrödinger equation. In this section, my aim is to introduce the various elements of the full resource theory of CTOs step by step, to ensure that its assumptions are transparent. In particular, starting with the evolution of a single, isolated quantum system, each of these steps will consist in allowing for a more general physical evolution by adding the possibility of the system to interact with additional systems. At each step, I will discuss the state transitions of the system that are possible at this step, and in this context also introduce the notion of a *monotone* with respect to the possible state transitions. This will set the stage for the later chapters, in which the resource theory of CTOs and the monotones will play a central role.

Consider, then, a finite-dimensional system  $S$  with an initial Hamiltonian  $H_S$  and an initial state  $\rho \equiv \rho(0)$ . According to the Schrödinger equation, the state  $\rho(t)$  of the system at a later time  $t$  is

$$\rho(t) = U(t)\rho(0)U(t)^\dagger,$$

where  $U(t) := \exp(-itH_S)$  is the unitary operator that is generated by  $H_S$ . In other words, the set of possible evolutions of  $S$  in isolation and starting from  $\rho$  is simply the set of unitary channels  $\mathcal{U}_t$  that are generated by  $H_S$  for all  $t \in \mathbb{R}$ . By virtue of the postulates of unitary quantum mechanics, if we wanted to be able to induce more general evolutions of  $S$ , then we have to allow for  $S$  to interact with additional systems. For instance, one can show that it is possible to realize *any* unitary that commutes with  $H_S$  (not just the one generated by the latter) by having  $S$  interact in an energy-preserving manner with an additional system. That is, let  $T$  denote a second quantum system with local Hamiltonian  $H_T$ . The joint Hamiltonian of  $S$  and  $T$  is

$$H_{ST} = H_S + H_T + H_{\text{int}},$$

where  $H_{\text{int}}$  is an interaction term that acts on  $ST$  and where here, and in the remainder of this thesis, I write  $H_A + H_B$  as short for  $H_S \otimes \mathbb{1}_B + \mathbb{1}_S \otimes H_B$ , whenever it is clear from the context that the involved Hamiltonians only have local support. We then have the following, where  $\text{Tr}_X$  denotes the partial trace operation over a subsystem  $X$  (which is the canonical map for subsystem states in quantum mechanics [39]):

**Lemma 1** ([40]). *The following are equivalent:*

1. There exists a unitary  $U$  such that  $[U, H_S] = 0$  and

$$\rho' = U\rho U^\dagger,$$

2. There exist  $H_T, H_{\text{int}}$  with  $[H_{\text{int}}, H_S \otimes \mathbb{1}_T] = 0$ , a  $t \in \mathbb{R}$  and an initial state  $\sigma$  on  $T$  such that

$$\rho' = \text{Tr}_T \left[ U_{ST}(t)(\rho \otimes \sigma)U_{ST}^\dagger(t) \right],$$

where  $U_{TS}(t) := \exp(-itH_{ST})$ .

In other words, by considering the additional system  $T$ , we enlarged the set of evolutions that we can induce on  $S$  from just those unitaries that are generated by  $H_S$  to any unitary that commutes with the latter. Note that, of course, we could also have generated energy-non-preserving unitaries on  $S$  by engineering an interaction  $H_{\text{int}}$  that does not preserve the energy. However, for the sake of conceptual clarity, we want to keep those ancillary systems that allow for the realization of general energy-preserving unitaries on  $S$  separate from those ancillary systems (i. e. batteries) that enable for energy to be pumped into and be extracted from  $S$ . As such, the value of Lemma 1 is that the authors of [40] provide an explicit construction for the  $H_T, H_{\text{int}}$  and  $\sigma$ .

### Thermal operations

Allowing for system  $T$ , we enlarged the set of possible evolutions on  $S$ , but different energy eigensubspaces of  $S$  cannot interact. In a next step, we allow for an additional system  $B$ , a heat bath of some fixed temperature  $\mathcal{T}$ , to also remedy this in a physically motivated fashion. In particular, we allow for  $B$  to have any Hamiltonian  $H_B$ , while constraining its initial state, for a given choice of  $H_B$ , to be the Gibbs state

$$\omega_\beta(H_B) := \frac{e^{-\beta H_B}}{Z(\beta, H_B)},$$

where  $Z(\beta, H) := \text{Tr}(e^{-\beta H})$  is the partition function with respect to  $\beta$  and some Hamiltonian  $H$  and  $\beta := 1/k_B \mathcal{T}$  is the inverse temperature,  $k_B$  being the Boltzmann constant. We set  $H_{SB} = H_S + H_B$  as the initial joint Hamiltonian of system, that is, bath and system are initially non-interacting. Now, we can apply Lemma 1 to apply any energy-preserving unitary  $U_{SB}$  on  $SB$  (in which  $SB$  now plays the role of  $S$  in the statement of the Lemma). Of course,  $U_{SB}$  will itself be generated according to the Schrödinger equation by a Hamiltonian  $H_{SB} + H_T + H_{\text{int}}$  that can have  $S$  and  $B$  interact arbitrarily strongly, but  $U_{SB}$  itself will always commute with  $H_{SB}$ .

The most general set of quantum channels on  $S$  that we can realize at this point are of the form

$$\mathcal{G}(\cdot) = \text{Tr}_B[U(\cdot \otimes \omega_\beta(H_B))U^\dagger], \quad (3.2.1)$$

where  $[U, H_{SB}] = 0$ . The set of quantum channels  $\mathcal{G}$  that can be written in the above form are called *thermal operations* [41, 42, 43]. In the following, we write  $\rho \rightarrow_{TO} \rho'$  if there exists a thermal operation  $\mathcal{G}$  such that  $\rho' = \mathcal{G}(\rho)$ .

### Resource theories and monotones

Thermal operations form an instance of a more general framework known as *resource theories*. The idea behind resource theories is to study a physical phenomenon of interest formally by

allowing systems to interact with ancillary systems in which the phenomenon of interest is absent and then study the resulting ordering over states that encodes which state transitions are possible. In the particular case of thermal operations, the phenomenon of interest is *thermal non-equilibrium*.<sup>2</sup> As such, we allow any system to interact with ancillary systems that are already in thermal equilibrium – Gibbs states with respect to arbitrary Hamiltonians – and to interact with them in such a way that the joint evolution itself cannot possibly increase the degree of thermal non-equilibrium – here represented by unitaries that commute with the total Hamiltonian of system and bath. In this way, the ordering  $\rightarrow_{TO}$  on states of  $S$  that is induced by the set of thermal operations is meant to formally represent those evolutions of  $S$  in the course of which the degree of thermal non-equilibrium can only decrease. In turn, only real-valued functions  $f$  on states of  $S$  that are *monotone* with respect to this ordering, that is, only functions for which

$$\rho \rightarrow_{TO} \rho' \Rightarrow f(\rho) \geq f(\rho'),$$

qualify, in principle, as a *measure* of thermal non-equilibrium (although there might be additional properties that one requires of useful measures such as continuity, additivity or faithfulness).

Resource theories have been studied abstractly [48, 49, 50, 51, 52] but there also exist dedicated resource theories for entanglement (better known as LOCC) [53, 54, 55], coherence [56, 57] (to which the author also has contributed in [5]), asymmetry [58, 59], magic state distillation [60], and several other physical phenomena. See [61] for a recent review. Apart from the variants of thermal operations that we will encounter in the following chapters, the use of the framework of resource theories also has been extended within quantum thermodynamics, where it has been generalized to thermodynamic potentials other than the Hamiltonian [62, 63] and even to non-commuting quantities [64, 65].

Before introducing the monotones for thermal operations and further widening the set of operations themselves, let me briefly comment on the assumption that a) the state of the bath is a Gibbs state, b) the initial state of system and bath is uncorrelated and c) that the Hamiltonian of the bath can be chosen freely. Concerning a), the motivation for this assumption is that, as we have seen, thermal operations are meant to model the most general way of having a system interact with a system in thermal equilibrium, and as such the choice of Gibbs states is due to them being the canonical state assigned to systems in thermal equilibrium. Of course, one may ask why this particular state is taken to represent systems in thermal equilibrium. This question will be the central topic of Chapter 6 and as such I refer the interested reader to this part of the thesis. Concerning b), from a pragmatic point of view, one can say that in the kinds of experiments and processes that quantum thermodynamics is meant to describe, the bath will generally still be much larger than the system  $S$ . As such, any correlations between the system and the bath that may be present initially would quickly disperse over the whole bath in such a way that any small subregion of the bath is essentially uncorrelated from the system. It should be noted, however, that this assumption introduces an asymmetry into the framework (product states evolving into correlated states) that makes thermal operations unsuitable to *explain* time asymmetry in thermodynamics. This is similar to the criticism of Boltzmann’s “Stoßzahlansatz” in the foundations of statistical mechanics. Finally, concerning c), we emphasize that thermal operations are meant to provide fundamental bounds to the kinds of thermodynamic evolutions that experimenters could engineer or observe in their

<sup>2</sup> While the framework of resource as I present it here is quite young, the idea to study thermodynamics in the above way goes back to the axiomatic thermodynamics of Lieb and Yngvason [44, 45] and even work of Carathéodory from 1909 [46]. See [47] for an investigation of the precise relation between axiomatic thermodynamics and thermal operations.

laboratories. This would be undermined if we restricted the kinds of Hamiltonians that we believe can be engineered in the lab *a priori*. A similar point can be made about the assumption that the joint system can evolve under any energy-preserving unitary. This of course assumes arbitrarily precise control over the interaction Hamiltonian  $H_{int}$  and is again motivated by not wanting to undermine the results of the framework by future developments. Indeed, a moment's thought reveals that the ability to fine-tune the evolution of joint system *implies* the ability to prepare any bath Hamiltonian. This is because, in a macroscopic heat bath, the energy levels are extremely dense, so that any Hamiltonian  $H_B$  can be "simulated" by a sufficiently skilled experimenter by having the system only interact with a subset of energy levels of the macroscopic bath that has the same gap and degeneracy structure. This is consistent with the framework of thermal operations because the thermal state of the macroscopic bath projected onto this truncated Hamiltonian  $H_B$  is still a thermal state, i.e. for any projector  $P$  onto a subset of energy levels of some Hamiltonian  $H$ , we have that [66]

$$\omega_\beta(PHP) = \frac{P\omega_\beta(H)P}{\text{Tr}(P\omega_\beta(H)P)}. \quad (3.2.2)$$

Now that we have extensively talked about the resource theory of thermal operations, let us focus on understanding its monotones. For this, it will be useful to define the notion of *d-majorization* as a relation between vectors [67].

**Definition 2** (d-majorization). *Let  $x, x', y \in \mathbb{R}^n$  with  $y$  strictly positive. Then we say that  $x$  d-majorizes  $x'$  with respect to  $y$ , denoted as  $d(x||y) \succeq d(x'||y)$ , if for any convex continuous function  $g$*

$$\sum_i y_i g\left(\frac{x_i}{y_i}\right) \geq \sum_i y_i g\left(\frac{x'_i}{y_i}\right).$$

Now, for a  $d$ -dimensional system with Hamiltonian  $H_S = \sum_i^d E_i |i\rangle\langle i|$  and in state  $\rho$ , define as  $P_{H_S}(\rho)$  the probability vector whose  $i$ th element is  $P_{H_S}(\rho)_i = \langle i|\rho|i\rangle$ , that is  $P_{H_S}(\rho)$  is the diagonal of  $\rho$  in the energy eigenbasis. We then write

$$d(\rho||\omega_\beta(H_S)) \equiv d(P_{H_S}(\rho)||P_{H_S}(\omega_\beta(H_S))).$$

The following theorem relates the notion of d-majorization to thermal operations:<sup>3</sup>

**Theorem 3** ([41]). *Let  $\rho, \rho'$  be quantum states with  $[\rho', \omega_\beta(H_S)] = 0$ . Then, the following are equivalent:*

1. *There exists a thermal operation  $\mathcal{G}$  such that  $\rho' = \mathcal{G}(\rho)$ ,*

2.

$$d(\rho||\omega_\beta(H_S)) \succeq d(\rho'||\omega_\beta(H_S)), \quad (3.2.3)$$

3.

$$d(\omega_\beta(H_S)||\rho) \succeq d(\omega_\beta(H_S)||\rho'). \quad (3.2.4)$$

Theorem 3 is significant at several levels: First of all we see that, by introducing baths, we have again significantly increased the set of state transitions that can be realized on  $S$ , in particular allowing weight to be shifted between different energy eigensubspaces. Secondly, while checking d-majorization may seem to be a computationally demanding task, in that the

<sup>3</sup> While [41] show Theorem 3 for the special case of thermal operations, it is in fact a simple corollary of more general theorems proven much earlier, dating back at least to [67] and [68].



definition involves an infinite number of functions, [41] show that Eq. (3.2.3) and Eq. (3.2.4) can in fact efficiently be checked by means a criterion called *thermo-majorization* (see also [67, 68, 69]). This means that the set of possible state transitions under thermal operations (at least for pairs of states that satisfy the conditions of Theorem 3) can be efficiently characterized. Thirdly, and most importantly in the current context, is that Theorem 3 establishes an (infinitely) large family of monotones under thermal operation, for the set of energy incoherent states. To see this, first let

$$\mathcal{S} = \{\rho \mid [\rho, H_S] = 0\}$$

denote the set of energy-incoherent states on  $S$ . Further, it is easy to see that thermal operations by definition can never produce coherences in the energy eigenbasis, that is, if  $\rho \in \mathcal{S}$ , then we also have  $\mathcal{G}(\rho) \in \mathcal{S}$  for any thermal operation  $\mathcal{G}$ . The reason is that the unitary  $U$  which generates  $\mathcal{G}$  according to Eq. (3.2.1) commutes with  $H_S + H_B$ , so that for any state  $\rho \in \mathcal{S}$ ,

$$\begin{aligned} [\rho \otimes \omega_\beta(H_B), H_{SB}] = 0 &\Rightarrow [U\rho \otimes \omega_\beta(H_B)U^\dagger, H_{SB}] = 0 \\ &\Rightarrow [\mathcal{G}(\rho), H_S] = 0, \end{aligned} \quad (3.2.5)$$

as claimed. Moreover, it is also straightforward to check that thermal operations have the thermal state  $\omega_\beta(H_S)$  as a fixed point. This can be seen, for example, by realizing that thermal states are, by definition, uniformly distributed over their energy subspaces (recall Eq. (3.2.2)). Hence, any unitary that commutes with  $H$  will act trivially on each of these subspaces, and hence on the whole state, so that

$$\begin{aligned} \mathcal{G}(\omega_\beta(H_S)) &= \text{Tr}_B(U\omega_\beta(H_S) \otimes \omega_\beta(H_B)U^\dagger) \\ &= \text{Tr}_B(U\omega_\beta(H_{SB})U^\dagger) \\ &= \text{Tr}_B(\omega_\beta(H_{SB})) = \omega_\beta(H_S). \end{aligned}$$

Combining the above two facts with Theorem 3 and the definition of d-majorization implies that, for fixed system Hamiltonian  $H_S$  and inverse temperature  $\beta$ , any continuous convex function  $g$  gives rise to two monotones  $f_g, h_g : \mathcal{S} \rightarrow \mathbb{R}$ , defined as

$$\begin{aligned} f_g : \rho &\mapsto \sum_i \omega_i g\left(\frac{p_i}{\omega_i}\right), \\ h_g : \rho &\mapsto \sum_i p_i g\left(\frac{\omega_i}{p_i}\right), \end{aligned}$$

where  $p_i \equiv P_{H_S}(\rho)_i$  and  $\omega_i \equiv P_{H_S}(\omega_\beta(H_S))_i$  (note that for energy-incoherent states  $P_{H_S}(\rho) = \text{mspec}(\rho)$ ).<sup>4</sup>

Each of these monotones establishes a necessary condition on the existence of a state transition under thermal operations. In the following, I will present some of these constraints and discuss in which sense one of them, the free energy difference (as defined for quantum systems), becomes the only relevant constraint in the thermodynamic limit, hence allowing us to derive the Second Law of phenomenological thermodynamics from thermal operations. We begin by discussing the (non-equilibrium) free energy. This corresponds to the choice  $g = -\log$ , in which case we have that  $h_{-\log}$  equals the (quantum) relative entropy

$$S(\rho \parallel \sigma) := \text{Tr}(\rho(\log \rho - \log \sigma))$$

<sup>4</sup> Similar reasoning lets us derive monotones for other resource theories as well. In particular, let  $\sigma$  be a fixed state of a set of operations defined in a resource theory. Then we can define monotones  $f_g$  and  $h_g$  for that resource theory where the Gibbs state is replaced by  $\sigma$ .

over  $\mathcal{S}$ . Now, it is easy to see that the relative entropy actually encodes the free energy difference of an incoherent non-equilibrium quantum state  $\rho \in \mathcal{S}$ . To see this, define

$$F(\rho) := \langle H_S \rangle_\rho - k_B T S(\rho)$$

to be the non-equilibrium free energy of a state  $\rho$ , where

$$S(\rho) := -\text{Tr}(\rho \log(\rho)) = -S(\rho \| \mathbb{1})$$

is the von Neumann entropy. Equipped with these new definitions, one finds that, for any  $\rho \in \mathcal{S}$ ,

$$\begin{aligned} \frac{1}{\beta} S(\rho \| \omega_\beta(H_S)) &= -\frac{1}{\beta} S(\rho) + \langle H \rangle_\rho + \frac{1}{\beta} \log(Z) \\ &= F(\rho) - F(\omega_\beta(H_S)), \end{aligned}$$

as claimed. Theorem 3 then implies that the decrease of the non-equilibrium free energy difference of a system in state  $\rho$  is a necessary consequence of any thermodynamic evolution modeled by thermal operations, but not sufficient in the *single-shot* regime, in which one only cares about the possible state-transitions that can be implemented for a single iteration of the process (or in which one only has access to a single copy of  $S$ .) What about other monotones? Returning to the choice of  $g = -\log$ ,  $f_g$  defines a monotone that is called the *vacancy* and that has recently been shown to be the dominant monotone not in the i.i.d. limit but in a setting in which part of  $S$  is used as a resource to cool other parts of it [42, 70]. These results show that the vacancy is intimately related to the Third Law of thermodynamics, exemplifying the scope of the framework of thermal operations and its ability to provide a conceptually transparent approach to the derivation of thermodynamics from the postulates of quantum mechanics. Many other choices of  $g$  correspond to monotones that are well-known and studied in the information theory literature. For instance, the class of monotones  $f_g$  for continuous convex functions  $g$  with  $g(1) = 0$  give rise to the so-called *f-divergences*, which include, among others, the Hellinger distance, the  $\chi^2$ -divergence, the variational divergence [71]. Hence, Theorem 3 again establishes that in all of these distances, the distance between any non-equilibrium state  $\rho$  and the thermal state  $\omega_\beta(H_S)$  can only decrease in the course of evolutions modeled by thermal operations.

### *The macroscopic limit*

We are now in a position to see how the Second Law of phenomenological thermodynamics emerges as a special case from thermal operations.<sup>5</sup> Let us introduce an additional piece of notation, writing  $\rho \xrightarrow{\epsilon}_{TO} \rho'$ , if there exists a  $\rho'_\epsilon \in \mathcal{B}_\epsilon(\rho')$  such that  $\rho \rightarrow_{TO} \rho'_\epsilon$ . Here,  $\mathcal{B}_\epsilon(\rho)$  denotes the  $\epsilon$ -ball around  $\rho$  in trace distance, that is,

$$\mathcal{B}_\epsilon(\rho) = \{\rho' \mid \frac{1}{2} \|\rho' - \rho\|_1 \leq \epsilon\}.$$

Then, it follows from standard results in information theory about typical sequences that for any pair  $\rho, \rho'$  of states on  $S$  with  $\rho' \in \mathcal{S}$  and such that  $S(\rho \| \omega_\beta(H_S)) > S(\rho' \| \omega_\beta(H_S))$  and any  $\epsilon > 0$ , there exists an  $n_\epsilon \in \mathbb{N}$  such that, for all  $n \geq n_\epsilon$ , [38]

$$\rho^{\otimes n} \xrightarrow{\epsilon}_{TO} (\rho')^{\otimes n}. \quad (3.2.6)$$

<sup>5</sup> For connections between thermal operations and the Third Law, see [42, 70, 72, 73]. For a resource theoretic approach to the First Law, see [74]

This shows that the free energy difference is the *only* relevant monotone for possible state transitions in the thermodynamic limit. As the free energy quantifies the work that can be extracted from a state (both phenomenologically as well as in quantum thermodynamics [75, 76, 77]), this reproduces the constraints on state transitions and work extraction to the ones that were known already to Helmholtz and his contemporaries.

At this point, we see how quantum thermodynamics is indebted, in its approach to the laws of thermodynamics, to the methods of information theory: One begins with the study of *single-shot* quantities, namely monotones that characterize the possible state transitions with respect to a given model of physical processes, for a single instance of the process. The thermodynamic limit then is simply given by the single-shot capacities in the special case that input and output states are i.i.d. As such, the framework in fact makes no assumptions about the underlying size of the systems, meaning that, in principle, the single quantum system may already be macroscopic.

Before moving on to introduce another central concept of this thesis, catalysis, it should be mentioned that one can use thermal operations to make much stronger statements about the thermodynamic limit than Eq. (3.2.6). In particular, the latter follows only as a special case of a more general result about *interconversion rates* under thermal operations. For two states  $\rho, \rho'$ , let  $R \equiv R(\rho \rightarrow \rho')$  denote the largest number such that, for any  $\epsilon > 0$ , there exists a sufficiently large  $n$  so that

$$\rho^{\otimes n} \xrightarrow{\epsilon} (\rho')^{\otimes (Rn)}.$$

Then it was shown in [76] that

$$R = \frac{S(\rho \| \omega_\beta(H_S))}{S(\rho' \| \omega_\beta(H_S))}.$$

This provides a generalization of Carnot's statement of the Second Law, according to which the efficiency of heat engines is fundamentally limited. See [78] for a full-blown resource theory of asymptotic interconversion and [79] for finite-size corrections to conversion rates.

It should also be mentioned that the i.i.d. limit is an idealization of the thermodynamic limit that is chosen mostly for mathematical convenience and analytical tractability. Real macroscopic systems are never actually described by a perfect product state. In reaction to this critical point, quantum thermodynamics has also made recent progress, with contributions made also by the present author, in analysing the thermodynamic limit for larger and physically more realistic classes of states than only i.i.d. states. One relevant result in this direction will be presented below, however, in order to be able to state it we first need to generalize the framework of thermal operations further.

### 3.3 CATALYSIS, EMBEZZLEMENT, COHERENCES

#### (Uncorrelated) catalysis

We now consider, on top of the heat bath and the system  $T$ , the possibility of a *catalyst*. As the name suggests, a catalyst is a system that may interact with  $S$  and  $B$  in the course of their joint evolution, but whose final state, at the end of the evolution, will have to be exactly identical to its initial state. An additional requirement is that, at the end of the interaction, the state of the catalyst is *uncorrelated* with the system (but not necessarily the bath). Formally we say that there exists a *catalytic thermal operation (CTO)* [38] that takes an initial state  $\rho$  to a final

state  $\rho'$  if, for any  $\epsilon > 0$ , there exists a (finite-dimensional) system  $C$  with Hamiltonian  $H_C$  and an initial state  $\sigma$  on  $C$  such that

$$\rho \otimes \sigma \rightarrow_{TO} \rho'_\epsilon \otimes \sigma, \quad (3.3.1)$$

where  $\rho'_\epsilon \in \mathcal{B}_\epsilon(\rho')$  and where the relevant Hamiltonian is

$$H_{SC} = H_S + H_C.$$

In this case, we write  $\rho \rightarrow_{CTO} \rho'$ . Physically, the role of catalysts is well motivated. In particular, a catalyst could be thought of as an engine or other devices that are employed to facilitate a thermodynamic evolution, but whose potential to be re-used for the facilitation of further state transitions on other systems remains uncompromised. Indeed, useful catalysts will usually be far from-equilibrium states (In turn, it is quite clear that thermal catalysts do not add any power over non-catalytic thermal operations).

Clearly, allowing for catalysts again greatly enlarges the set of possible state transitions that can be realized on  $S$ . Indeed, one can neatly characterize the possible state transitions under uncorrelated catalysis in terms of a one-parameter family of functions that form a strict subset of the monotones for thermal operations. These are the *Rényi divergences*, which were first introduced [80] (see [81] for various quantum generalizations). For general quantum states  $\rho$  and  $\sigma$ , these are defined as

$$R_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log(\text{Tr}(\rho^\alpha \sigma^{1-\alpha}))$$

for  $\alpha \in (-\infty, 1) \cup (1, \infty)$  and as the corresponding continuous extension

$$R_\alpha(\rho\|\sigma) := \lim_{\gamma \rightarrow \alpha} R_\gamma(\rho\|\sigma)$$

for  $\alpha \in \{-\infty, 1, \infty\}$ . The Rényi divergences arise as  $d$ -majorization monotones for thermal operations as

$$F_\alpha(\rho) := R_\alpha(\rho\|\omega_\beta(H_S)) = g_\alpha^{-1}(h_{g_\alpha}(\rho)),$$

where  $g_\alpha(x) = \exp((1 - \alpha)x)$  and  $\rho \in \mathcal{S}$  [71]. Here, we define  $F_\alpha$  as generalized free energy differences (dropping the implicit dependence on  $\beta$ ). Since  $g_\alpha$  are strictly monotone, it is clear that  $F_\alpha$  inherit their monotonicity from that of  $h_g$ . Just like the  $f$ -divergences, the family of Rényi divergences is related to several well-studied quantities in information theory. To begin with, it is easy to see using l'Hopital's rule that  $R_1(\rho\|\sigma) = S(\rho\|\sigma)$ , that is, the relative entropy is one of the Rényi divergences. Other notable cases include the hypothesis testing relative entropy [82] for  $\alpha = 0$ , the Bhattacharyya distance for  $\alpha = 1/2$  [83] and (the logarithm of) the expected ratio  $\langle \frac{P_{H_S}(\rho)}{P_{H_S}(\sigma)} \rangle_{P_{H_S}(\rho)}$  for states  $\rho, \sigma \in \mathcal{S}$  for  $\alpha = 2$ . In [38], the following was shown:

**Theorem 4** (Second Laws). *Let  $\rho, \rho' \in \mathcal{S}$ . Then the following are equivalent:*

1.  $\rho \rightarrow_{CTO} \rho'$
2.  $F_\alpha(\rho) \geq F_\alpha(\rho'), \quad \alpha \in [-\infty, \infty]$

In other words, by allowing for a catalytic ancilla many of the monotones under thermal operations stop being monotones, further enlarging the set of possible state transitions. Indeed, one can further get rid of the constraints corresponding to negative  $\alpha$  by allowing for a

further ancilla  $C'$ , a single qubit initially in the state  $|0\rangle\langle 0|$  that has to be returned arbitrarily closely to its initial state. That is, let us write  $\rho \xrightarrow{\epsilon}_{CTO'} \rho'$  if

$$\rho \otimes |0\rangle\langle 0| \xrightarrow{\epsilon}_{CTO} \rho' \otimes |0\rangle\langle 0|$$

and write  $\rho \rightarrow_{CTO'} \rho'$  if  $\rho \xrightarrow{\epsilon}_{CTO} \rho'$  for any  $\epsilon > 0$ . For this further weakening of thermal operations, one can show a stronger version of Theorem 4 for pairs of states  $(\rho, \rho')$  such that  $[\rho, \sigma] = [\rho', \sigma] = 0$ ,  $\rho \rightarrow_{CTO} \rho'$  if and only if, for all  $\alpha \geq 0$ ,

$$F_\alpha(\rho) \geq F_\alpha(\rho'). \quad (3.3.2)$$

The above characterization of thermal operations under catalysis also allows for elegant statements that describe the approximation of the macroscopic Second Law not only in the asymptotic limit, as around Eq. 3.2.6, but also for finitely many copies of a state. To state one particularly simple such bound, define, for any state  $\sigma$ ,

$$V(\rho\|\sigma) := \text{Tr} \left[ \rho \left( \log \left( \frac{\rho}{\sigma} \right) - S(\rho\|\sigma) \right)^2 \right]$$

as the *relative variance of information* [84, 85, 86], where  $\log \left( \frac{\rho}{\sigma} \right) \equiv \log(\rho) - \log(\sigma)$  denotes the *relative surprisal*, and let  $V_\beta(\rho) := V(\beta\|\omega_\beta(H_S))$ .  $V_\beta$  describes, in a sense, the fluctuations of  $\rho$  around the free energy difference. Recent and yet unpublished work co-authored by the present author [8] then implies the following:

**Theorem 5.** *For fixed,  $\beta, H_S$ , let  $(\rho, \rho') \in \mathcal{S}^2$  be a pair of energy-incoherent states. Then, for any  $k > 0$ , there exist  $\epsilon_1 \geq 0$  and  $\epsilon_2 \geq 0$  with  $\epsilon_1 + \epsilon_2 \leq 1/k^2$  and states  $\rho_\epsilon \in \mathcal{B}_\epsilon(\rho), \rho'_\epsilon \in \mathcal{B}_\epsilon(\rho')$  such that*

$$\rho_\epsilon \rightarrow_{CTO} \rho \quad \text{and} \quad \rho'_\epsilon \rightarrow_{CTO} \rho'_\epsilon$$

and

$$F_\alpha(\rho_\epsilon) \geq F_1(\rho) - k\sqrt{V_\beta(\rho)}, \quad \alpha \in [0, 1]$$

$$F_1(\rho'_\epsilon) \geq F_\alpha(\rho'_\epsilon) - k\sqrt{V_\beta(\rho')}, \quad \alpha \geq 1.$$

This theorem connects the Rényi divergences for  $\alpha \geq 0$  to the variance of surprisal of the respective initial and final states. By combining this statement with Eq. (3.3.2), we can then infer the following corollary to the above theorem:

**Corollary 6** (Single-shot sufficient conditions for  $\succ_{\sigma, \epsilon}^C$ ). *For fixed  $\beta, H_S$ , let  $(\rho, \rho') \in \mathcal{S}^2$  be a pair of energy-incoherent states such that*

$$F(\rho) - F(\rho') =: \delta > 0$$

For any  $n \in \mathbb{N}$ , we have

$$\rho^{\otimes n} \xrightarrow{\epsilon}_{CTO'} (\rho')^{\otimes n}$$

with

$$\epsilon \leq \frac{4(V(\rho\|\sigma) + V(\rho'\|\sigma))}{n\delta^2}.$$

While this result clearly reproduces the macroscopic Second Law in the limit of  $n \rightarrow \infty$ , it also produces simple bounds for finite  $n$ . Moreover, it implies that state transitions between states whose variance  $V(\rho\|\sigma)$  is negligible already at the single-shot level are also essentially governed by the free energy alone. One example of this are Gibbs states of an infinite lattice with local Hamiltonians and at sufficiently high temperature, and more generally ergodic

states, i.e. translation-invariant lattice states that have vanishing fluctuations in the statistics of any local observable that is averaged over some region of the lattice, in the limit of this region becoming infinitely large [87, 88]. As such, this result shows that the resource theoretic approach of quantum thermodynamics can also be useful for considering thermodynamic limits that go beyond the simple i.i.d. limit described earlier. See [79, 89] for similar discussions.

### *Embezzlement and approximate catalysis*

Before moving on, let me briefly comment on the requirement to have the catalyst returned exactly, instead of only approximately. That is, instead of defining CTOs via Eq. (3.3.1), it might seem operationally more reasonable to require the weaker condition that

$$\rho \otimes \sigma \xrightarrow{\epsilon}_{TO} \rho' \otimes \sigma. \quad (3.3.3)$$

However, if no additional requirements on the scaling of the dimension of the catalyst are made (such as in the case of CTO's discussed in the previous section), then this alternative definition of CTOs becomes trivial, in the sense that  $\rho \rightarrow_{CTO} \rho$  for any  $(\rho, \rho') \in \mathcal{S}^2$ . This phenomenon is known as *embezzling*, named after a similar phenomenon in the resource theory of entanglement [90]. The reason for the trivialization is essentially that, by making the catalyst large enough, we can “hide” any difference between the initial and the final state in the catalyst, making the trace distance as small as we like. Indeed, a simple way to construct such an embezzling catalyst for a given transition  $\rho \rightarrow \rho'$  is to choose  $\sigma_n = (\rho')^{\otimes n}$  for some  $n \in \mathbb{N}$ . Then, we can use results from [91, 92] to show that

$$\rho \otimes \sigma_n \rightarrow_{TO} \tau,$$

where

$$\|\tau - \rho' \otimes \sigma_n\|_1^2 \leq \frac{1}{2} \log\left(1 + \frac{K}{n}\right)$$

for some constant  $K$  that depends on  $\rho$  and  $\rho'$  but not  $n$ . Clearly, this implies that any transition would be possible using a catalyst under the modified definition Eq. (3.3.3). Conversely, once can show that when imposing restrictions on the size of the catalyst, then trivialization in the above sense does not occur. Indeed, results in [38] imply that the dimension of the catalyst bounds the extent to which the free energy can be reduced under “modified” CTOs (in the sense of Eq. (3.3.3)).

### *Thermal operations in the presence of coherences and GP-maps*

Most of the above results only hold if the final state  $\rho'$  of a state transitions has no coherences in the energy eigenbasis. This is mostly because characterizing those state transitions is significantly simpler, allowing for the application of standard results from the theory of d-majorization for vectors, as presented on the preceding pages. However, researchers have also put significant effort in studying the more complicated case in which energy coherences are present in both initial and final states. While these results do not matter much for results presented in this thesis, let me here present a brief overview over the main findings: In [38, 93], a number of necessary conditions for the existence of a CTO for a generic transition  $\rho \rightarrow_{CTO} \rho'$  was given in terms of Rényi divergences. In [94, 95], it was shown that energy coherences can be understood as a resource independently of the “non-equilibriumness” captured by the d-majorization monotones presented above. Indeed, later it was shown that these

coherences could be used to extract work [96, 97] (see also [98] for a review). Recently, [99] finally provided fully general necessary and sufficient conditions for state transitions under thermal operations. They did so by using more general results based on a generalization of d-majorization to quantum systems, which they call *quantum majorization*. The details of quantum majorization, which is closely related to the theory of degradable channels [100, 101] and quantum dichotomies [102], lie beyond the scope of this thesis.

Another interesting subject-matter related to coherence, which will not be discussed in further detail in this thesis, is the relationship between thermal operations and the class of channels  $\tilde{\mathcal{G}}$  whose defining property is that, for a system with fixed system Hamiltonian  $H_S$  and bath inverse temperature  $\beta$ , they preserve the Gibbs state,

$$\tilde{\mathcal{G}}(\omega_\beta(H_S)) = \omega_\beta(H_S).$$

These channels are called *Gibbs-preserving (GP)-maps* and, as we have seen by the argument above Eq. (3.2.5), thermal operations form a subset of GP-maps. Indeed, and importantly, they only form a *strict* subset of GP-maps.<sup>6</sup> This separation leads to significant differences in the kinds of state transitions that can be realized by them. In particular, in [108], it was shown that there exist GP-maps that map an initially incoherent state to a coherent state. While any GP-map, by virtue of being a quantum channel, can be understood as the effective map that results from the interaction with another system (via Stinespring's dilation theorem [39]), the preceding sections do not provide us with a detailed and systematic understanding of these interactions (for GP-maps that are not thermal operations) and we will not consider them further.

#### *Further extensions of thermal operations*

On the preceding pages, we have seen how the framework of resource theories provides us with a conceptually transparent path from the postulates of quantum mechanics to a theory of the thermodynamic evolution of quantum systems that both implies (some of) the laws of phenomenological thermodynamics as special cases but also extends those laws from the thermodynamic limit into the realm of single-shot thermodynamics, which is governed by small systems, finite-size baths, fluctuations that are of the same order of magnitude as the size of the system, and single or few copies of the initial state. Researchers have extensively and fruitfully studied various further additions to the framework of CTOs, such as for example, different models of batteries to study work extraction and work fluctuations [75, 77, 109, 110, 111, 112], sources of coherences [113], time-dependent Hamiltonians [76] and various generalizations of catalysis [114, 115, 116]. Each of these extensions was operationally motivated to allow for the study of one aspect of the thermodynamic evolution, in an attempt to keep the various quantities and resources that affect quantum thermodynamical processes conceptually as cleanly separated as possible. Many of the results presented in the following chapters will also be based on particular modifications or extensions of CTOs.

<sup>6</sup> In Chapter 4, we will see that thermal operations and GP-maps are generalizations of channels that, in the mathematical physics literature, are called exactly factorizable maps and unital maps. Here, the fact that exactly factorizable maps form a strict subset of unital channels is closely related to the fact the seminal Birkhoff-von Neumann theorem on the decomposition of doubly stochastic matrices into a convex sum of permutations does not exist for quantum channels. See [103, 104, 105, 106, 107] for literature on the relation between unital and exactly factorizable maps.





## CATALYTIC QUANTUM RANDOMNESS

---

The work and results presented in this chapter deal primarily with the notion of randomness as a resource. In particular, we define models of random processes in which a system interacts with a source of randomness in either a classical or a quantum way and prove that these models differ in strength, in the sense that one and the same random transition of a system can be realized with a strictly smaller source of randomness if the process is quantum as compared to classical. This is interesting in its own right but can also be used in various applications ranging from decoherence [117] to private quantum channels [118] that are presented below.

That randomness and thermodynamics are intimately connected has been known at least since the work of Maxwell and Boltzmann, who put random distributions over particle velocities in a gas at the very beginning of their attempts to derive the laws of thermodynamics [19]. However, the models of random processes that underlie this chapter can be related to the theory of thermal operations as it was presented in chapter 3 in a very precise and formal manner. The main purpose of this introduction is to clarify this formal connection and spell it out.

### 4.1 RANDOM UNITARY CHANNELS AND THERMAL OPERATIONS

In quantum information theory, the most general description of a physical process is a quantum channel, which is a completely positive trace-preserving (CPTP) map [39]. We say that a given channel  $\mathcal{C}$  describes a (non-trivial) random process if it can be decomposed into a (non-trivial) convex combination of other CPTP channels, that is,

$$\mathcal{C} = \sum_i^k p_i \mathcal{C}_i, \quad (4.1.1)$$

where  $p_i > 0$  and  $\sum_i^k p_i = 1$ . Random processes can be thought of as an agent sampling from a classical random variable  $X$  with  $\text{Prob}(X = i) = p_i$  and then implementing the channel  $\mathcal{C}_i$  depending on the occurrence of event  $i$ . Of course, such a decomposition is in general not unique and some decompositions are more interesting than others. For instance, when one is interested in “separating” the randomness in a process from other aspects of it, it makes sense to ask for decompositions in which every channel  $\mathcal{C}_i$  is itself no random process. Moreover, when we are interested in connecting randomness with the notion of *irreversibility*, as we are in thermodynamics, then it makes sense to further ask for every channel  $\mathcal{C}_i$  to be reversible, that is, there should exist channels  $\mathcal{D}_i$  such that, for every state  $\rho$  and  $i$ ,

$$\mathcal{D}_i \circ \mathcal{C}_i(\rho) = \rho.$$

But the only reversible CPTP channels are unitary channels [119] and so random processes that admit such a decomposition are known as *random unitary channels (RUC)* [120].<sup>1</sup>

Random unitary channels can be linked to thermal operations in several different ways. To begin with, consider the frameworks of both thermal operations and catalytic thermal

<sup>1</sup> More generally, one can also consider convex mixtures of isometries between different spaces. The essentials of what the presentation in this section carry over to this more general case, but since it won't be required to understand the results, I here focus on the simpler case of random unitary channels.

operations but with all the Hamiltonians being fully degenerate (i.e.  $H_S \propto H_B \propto H_C \propto \mathbb{1}$ ). In this case, the commutation constraint on the unitary,  $[U, H_S + H_B + H_C] = 0$ , in the definition of CTOs trivializes regardless of the value of  $\beta$ , and the thermal states are maximally mixed. In other words, in this setting, “thermal” operations then are the set of channels  $\mathcal{G}$  that can be written as

$$G(\cdot) = \text{Tr}_B \left[ U(\cdot \otimes \mathbf{1}_m) U^\dagger \right], \quad (4.1.2)$$

where  $m \in \mathbb{N}$ ,  $\mathbf{1}_m \equiv \mathbb{1}/m$  denotes the maximally mixed state on an  $m$ -dimensional Hilbert space  $\mathcal{H}_B$ , and  $U$  is any unitary acting on  $\mathcal{H}_S \otimes \mathcal{H}_B$ . This set of channels are known as *noisy operations* (NO) [121] and they bear the following simple relation to random unitary channels: Given two states  $\rho, \rho'$  on  $S$ , there exists a noisy operation  $\mathcal{G}$  such that  $\mathcal{G}(\rho) = \rho'$  if and only if there exists a random unitary channel  $\mathcal{R}$  with (possibly irrational) weights  $\{p_i\}_i^k$  such that  $\mathcal{R}(\rho) = \sum_i p_i \mathcal{U}_i(\rho) \in \mathcal{B}_\epsilon(\rho')$ . For the case of rational weights  $\{p_i\}$ , this is particularly simple to see in one direction: Let  $\{p_i = a_i/l, \mathcal{U}_i\}_i^k$  with  $a_i, l \in \mathbb{N}$  be some decomposition of a random unitary channel  $\mathcal{R}$  into rational weights and unitary channels  $\mathcal{U}_i(\cdot) = U_i \cdot U_i^\dagger$ . Then choose  $m = l$  and define the unitary

$$U = \sum_i^k \sum_{j=1}^{a_i} U_i \otimes |i, j\rangle\langle i, j|, \quad (4.1.3)$$

where  $\{|i, j\rangle\}$  is an orthonormal basis for  $\mathcal{H}_B$ . It is easy to check that the corresponding noisy operation  $\mathcal{G} = \mathcal{R}$ , so that they coincide on all state transitions, establishing the claim. In other words, when the unitary defined in Eq. (4.1.3) is applied to system and bath, Eq. (4.1.2), the maximally mixed state on  $B$  serves as the randomness that an agent uses to decide which unitary channel  $\mathcal{U}_i$  to apply to  $S$ . Roughly speaking, then, this neat equivalence between random unitary channels and noisy operations (at the level of possible state transitions, not in general (of channels) shows that, in the special case in which all energetic considerations (encoded in the Hamiltonians of bath and system) are absent, thermal operations reduce to a model of random processes, in which the heat bath acts as the sole source of randomness. Of course, all of these considerations apply just as well in the presence of catalysts, in which case CTOs reduce to *catalytic noisy operations* (CNO).

While the above correspondence will concern us in more detail below, let me emphasize that random unitary channels are not only related to thermal operations in the case of degenerate Hamiltonians. Instead, any transition  $\rho \rightarrow_{TO} \rho'$  that is possible using thermal operations can be realized in a two step-process: To do so, we split the bath  $B$  into two subregions,  $\tilde{B}$  and an  $m$ -dimensional “source of randomness”  $R$ , with joint Hamiltonian

$$H_B = H_{\tilde{B}} \otimes \mathbb{1}_R.$$

Now, first we map  $\rho$  to the joint product state  $\rho \mapsto \rho \otimes \omega_\beta(H_{\tilde{B}})$ . In a second step, we then apply a series of random unitary channels on  $S\tilde{B}$ , where each of these random unitary channels has non-trivial support only *within* a single eigensubspace of the Hamiltonian  $H_S + H_{\tilde{B}}$ . This two-step process can be realized as a thermal operation by means of the correspondence between random unitary channels and noisy operations sketched above and the construction Eq. (4.1.3). While formally trivial, the above construction proves conceptually insightful, because it shows that we can formally distinguish two contributions of the heat bath in a thermodynamic process described by thermal operations: The bath contributes both randomness and acts as an energetic heat reservoir (whose contribution is represented by the non-trivial Hamiltonian  $H_{\tilde{B}}$ ). Again, an analogous statement holds for CNOs.

## 4.2 MAJORIZATION

The above connection between thermal operations and random processes also straightforwardly translates to the level of monotones. In particular, setting  $H_S \propto H_B \propto \mathbb{1}$ , Theorem 3 implies that a sufficient and necessary conditions for the existence of a noisy operations between two states  $\rho = \sum_i p_i |i\rangle\langle i|$  and  $\rho' = \sum_j p'_j |j\rangle\langle j|$  is that, for any continuous convex function  $g$ ,

$$\sum_i g(p_i) \leq \sum_j g(p'_j). \quad (4.2.1)$$

This condition, which can be defined for any pair of real vectors is known as *majorization* [69] and denoted as  $x \succeq y$  for any pair of vectors  $x, y$ . We can easily extend it to a preorder<sup>2</sup> on density matrices, where we write  $\rho \succeq \rho'$  to mean that  $\text{spec}(\rho) \succeq \text{spec}(\rho')$ . Note that since for trivial Hamiltonians there is no preferred energy basis anymore, for the majorization relation the conditions Eq. (4.2.1) are already fully quantum and the complications surveyed in Sec. 3.3 do not arise.

As is implied by the previous discussion, majorization characterizes state transitions that can be realized by random unitary processes and noisy operations.<sup>3</sup> However, from Eq. (4.2.1), it might not be apparent in what sense majorization orders vectors by how “random” they are. Luckily, there exists an equivalent condition for majorization (for the case in which both vectors have the same dimension) from which the connection between majorization and randomness becomes more apparent. Namely, for two vectors  $x, y \in \mathbb{R}^d$ , let  $x^\downarrow, y^\downarrow \in \mathbb{R}^d$  denote the vectors obtained by reordering  $x$  and  $y$  respectively non-increasingly, so that  $x_i^\downarrow \geq x_{i+1}^\downarrow$  for all  $i \in \{1, \dots, d-1\}$  and similarly for  $y^\downarrow$ . Then we have that  $x \succeq y$  if and only if,

$$\begin{aligned} \sum_{i=1}^k x_i^\downarrow &\geq \sum_{i=1}^k y_i^\downarrow, \quad k \in \{1, \dots, d-1\} \\ \sum_{i=1}^d x_i^\downarrow &= \sum_{i=1}^d y_i^\downarrow. \end{aligned}$$

This condition provides a better intuitive link between majorization and randomness than Eq. (4.2.1), showing that majorization is essentially a measure of how evenly the weight in a vector is distributed: Random processes described by noisy operations and random unitary channels spread the weight of a probability vector across the whole state. Indeed, for quantum states (for which the second condition holds true by definition), we readily see that pure states are the maxima of the majorization ordering, while, for a fixed dimension  $d$ , the maximally mixed state  $\mathbb{1}/d$  is the unique minimum.

Monotones under majorization are called *Schur-convex* functions. All of the monotones under thermal operations yield Schur-convex functions if we set  $\omega_\beta(H_S) = \mathbb{1}/d_S$ , where  $d_S$  is the dimension of the system. Moreover, everything carries over to the case of catalysis, where we find that a transition  $\rho \rightarrow_{\text{CNO}} \rho'$  is possible by means of an additional catalyst if and only if, for all  $\alpha \in (-\infty, \infty)$ ,

$$R_\alpha(\rho) := R_\alpha(\rho \| \mathbb{1}) \leq R_\alpha(\rho'). \quad (4.2.2)$$

The relation between states transferrable under CNOs is also known as *trumping* and the conditions Eq. (4.2.2) were first shown in [125, 126]. The functions  $R_\alpha$  are known as *Rényi*

<sup>2</sup> A preorder is weaker than a partial order in that it is a binary relation that is transitive and reflexive but not necessarily anti-symmetric.

<sup>3</sup> Majorization has also been found to characterize the state transitions for pure bipartite states in the resource theory of entanglement [122, 123] and to be related to the workings of several well-known quantum algorithms [124].

entropies and just like in the case of the Rényi divergences, we find that  $R_1 = S$ , that is, the special case  $\alpha = 1$  corresponds to the von Neumann entropy.

### 4.3 RANDOMNESS AS A RESOURCE

With the relationship between randomness, as modeled by noisy operations and random unitary channels, and thermal operations clarified, we now turn to the work presented in this chapter. In introducing the resource theory of thermal operations, we have assumed full control on the side of an experimenter at both the level of the bath Hamiltonians that can be prepared, as well as at the level of the unitaries that can be implemented and the catalyst states that can be prepared. This let us derive fundamental bounds on possible state transitions. However, from a practical point of view, it is just as interesting to study how the set of realisable state transitions changes if the degree of control is limited. Several works address this question, for example by focusing on finite-size baths [7, 127, 128] or control restrictions [129, 130, 131]. In particular, one of the works co-authored by the present author, [7], investigates bounds on the size of a heat bath required to thermalize systems that exhibit a phenomenon known as many-body localization [28, 132]. Along similar lines, the central idea in the following is to place limits on the amount of randomness that is available to an agent, that is, to study randomness as a resource. We have seen how transitions under thermal operations can be implemented as a sequence of random unitary channels on the energy eigensubspaces of a joint system  $S\tilde{B}$  and that the randomness required to implement these random unitary channels derives from a part  $R$  of the bath that acts as a source of randomness. But what if this part  $R$  is small? Or what if the interaction between system  $S$  and bath  $B$  only lasts a short time so that the effective size of the region  $R$  with which  $S$  can interact in this time is small? In this case intuitively there is less randomness “available” to implement the random unitary channels and so a natural question to ask is: If the size of  $R$  is constrained, what are the state transitions under noisy operations and random unitary channels that can be realized? This is the question that we answer in this chapter, showing that there exists a separation between the transitions that can be realized with noisy operations and random unitary channels respectively.

To ask for the state transitions that can be implemented with a fixed-size source of randomness is closely related to questions about the least number of unitaries such that randomly sampling from them produces a desired effect. To see this, note that we could think of the fixed-size source of randomness as a limited number  $n$  of coin flips that I can make to decide which unitary to implement on  $S$ . To ask whether I can realize a desired transition  $\rho \succeq \rho'$  using only those coin flips then is equivalent to asking whether there exists a random unitary channel  $\mathcal{C}$  such that  $\rho' = \mathcal{C}(\rho)$  and for which there exists a decomposition of the form in Eq. (4.1.1) into unitary channels with  $k \leq 2^n$ .<sup>4</sup> This kind of question has been investigated in several ways already. For example, in a seminal work [134], the authors were able to operationally characterize the quantum mutual information of a bipartite quantum state as the asymptotic number  $n$  of coin flips required to implement a random unitary channel that decouples the state, that is, destroys all the correlations between the margins. Similar ways to study randomness as a resource can be found in [92, 135]. In a similar vein, researchers have been interested in the number of unitaries that are required to approximate a dephasing channel [136] or the effect of a Haar random unitary channel [137, 138]. All of these can be interpreted as investigations about the most efficient use of a limited amount of available

<sup>4</sup> Moreover, results in [107, 133] imply that in considering this question, one can assume without loss of generality that the distribution over unitaries is uniform, i.e.  $p_i = 1/k$  for all  $i$ .

randomness. In the following publication [1], we study this question in full generality and also distinguish between the power of random unitary channels, in which  $R$  is implicit and which provide a classical model of randomness, and noisy operations, in which  $R$  is explicit and which therefore provides a quantum model of randomness. The separation results that we prove add to previous research showing that using quantum systems to implement a random process provides possibilities that are absent for classical systems [139] and provide a kind of quantum advantage that is quite different in nature to those studied in quantum computation.

As an additional result, we present an expander graph construction based on the results in [138, 140, 141] that yields a dephasing map that dephases a system exponentially quickly, in 2-norm, in the size of the the source of randomness<sup>5</sup>. This result complements the above results in that it provides a study of the required randomness for stochastic processes in a different norm and for situations in which the source of randomness is much smaller than the system.

---

<sup>5</sup> We note that there is a typo in the statement of the corresponding Theorem 3, which misstates the convergence as being exponential in the *dimension* of the source of randomness. An erratum has been submitted to PRX.



## Catalytic Quantum Randomness

P. Boes,<sup>1</sup> H. Wilming,<sup>1,2</sup> R. Gallego,<sup>1</sup> and J. Eisert<sup>1</sup>

<sup>1</sup>*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*

<sup>2</sup>*Institute for Theoretical Physics, ETH Zürich, 8093 Zürich, Switzerland*



(Received 13 June 2018; revised manuscript received 21 August 2018; published 29 October 2018)

Randomness is a defining element of mixing processes in nature and an essential ingredient to many protocols in quantum information. In this work, we investigate how much randomness is required to transform a given quantum state into another one. Specifically, we ask whether there is a gap between the power of a classical source of randomness compared to that of a quantum one. We provide a complete answer to these questions, by identifying provably optimal protocols for both classical and quantum sources of randomness, based on a dephasing construction. We find that in order to implement any noisy transition on a  $d$ -dimensional quantum system it is necessary and sufficient to have a quantum source of randomness of dimension  $\sqrt{d}$  or a classical one of dimension  $d$ . Interestingly, coherences provided by quantum states in a source of randomness offer a quadratic advantage. The process we construct has the additional features to be robust and catalytic; i.e., the source of randomness can be reused. Building upon this formal framework, we illustrate that this dephasing construction can serve as a useful primitive in both equilibration and quantum information theory: We discuss applications describing the smallest measurement device, capturing the smallest equilibrating environment allowed by quantum mechanics, or forming the basis for a cryptographic private quantum channel. We complement the exact analysis with a discussion of approximate protocols based on quantum expanders deriving from discrete Weyl systems. This gives rise to equilibrating environments of remarkably small dimension. Our results highlight the curious feature of randomness that residual correlations and dimension can be traded against each other.

DOI: [10.1103/PhysRevX.8.041016](https://doi.org/10.1103/PhysRevX.8.041016)

Subject Areas: Quantum Information

### I. INTRODUCTION

Randomness is a central concept and resource in various fields of research in computer science, information theory, and physics, in both the classical and the quantum realm. It is an ingredient to (quantum) algorithm design, a core element in coding and communication protocols, and plays a central role in fundamental aspects of statistical mechanics. In the quantum context, randomness is also increasingly being seen as a valuable resource. A natural question that arises in this context is then how much of it is required to implement a given physical process on a quantum system. Another important question is to what extent the required amount of randomness differs depending on whether an *implicit* or an *explicit* model of randomness is employed. Here, an implicit model of randomness considers the source of randomness (SOR) as a black box that provides coin flips, while an explicit model takes into account the fact that, fundamentally, all systems including the ones provided by

the SOR are quantum systems, and hence models the randomness as a quantum state.

In this work, we give a complete answer to both of the above questions. We provide, for both the implicit and explicit model, optimal and tight bounds on the amount of randomness required to implement physical processes on quantum systems. Moreover, we show a strict separation between the above models, in the sense that every physical process can be implemented in the explicit model by using only half the amount of randomness that is required in the implicit model.

Specifically, we use a model of noisy processes—processes that require randomness—known as noisy operations [1]. We study the minimal amount of noise required to implement a large variety of noisy processes and construct protocols that saturate the lower bounds imposed by quantum mechanics. These processes include dephasing and equilibration [2,3], decoherence [4,5], the implementation of measurements [5–7], any transition between two quantum states that requires randomness [1], as well as the novel construction of private quantum channels [8,9].

It is an important aspect of our work that, by virtue of an explicit model, these saturated lower bounds also translate into bounds on the physical size of a SOR. This insight allows us to construct, for particular processes, the smallest decohering

---

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.*

environment or measurement device compatible with quantum mechanics [4]. In other words, it provides an understanding of the smallest equilibrating environment [2] possible. The surprisingly small size that suffices for an environment to be equilibrating challenges the commonly held view that such decohering baths should necessarily feature a large dimension.

A further notable feature of the protocols that we construct is that they are catalytic: The same unit of randomness can be reused for different processes [10]. It is also robust, in the sense that we do not require perfect control in either the states prepared by the SOR or the timing of the process, and further recurrent, in the sense that, for large system dimension  $d$ , continuous time versions of our noisy processes maintain a state close to the desired final state for times  $\tau \propto \sqrt{d}$ , at which point the system recurs to the initial state.

## II. CLASSICAL VERSUS QUANTUM NOISE

Let us begin with discussing in more detail the difference between classical and quantum uses of randomness. Consider initial and final (mixed) states  $\rho, \rho'$  on a Hilbert space  $\mathcal{H}_S$  of dimension  $\dim(\mathcal{H}_S) = d$ . We are concerned with the possibility of implementing a transition  $\mathcal{E}(\rho) = \rho'$ , where  $\mathcal{E}$  represents a noisy process. There exist different ways of modeling the maps  $\mathcal{E}$ , which we now explain in detail.

In a classical, implicit model of the SOR one assumes a discrete random variable  $J$  that is uniformly distributed over  $m$  possible values. Depending on the value of  $j$ , one implements a given unitary transformation  $U_j$ , which gives rise to the operations

$$\mathcal{E}_C^m(\cdot) = \frac{1}{m} \sum_{i=1}^m U_i \cdot U_i^\dagger. \quad (1)$$

If there exist  $\mathcal{E}_C^m$  so that a transition is possible, we simply denote it by  $\rho \xrightarrow{m} \rho'$ . In contrast, in an explicit quantum model, the SOR is a quantum system  $R$  in the maximally mixed state of dimension  $m$ , which we denote by  $\mathbb{I}_m := (1/m)\mathbb{1}$ , with  $\mathbb{1}$  being the identity matrix. In this model, noisy processes are any effect of a unitary joint evolution of the compound,

$$\mathcal{E}_Q^m(\cdot) = \text{tr}_R[U(\cdot \otimes \mathbb{I}_m)U^\dagger]. \quad (2)$$

As in the classical case, we write  $\rho \xrightarrow{m} \rho'$  whenever the transition is possible.

The set of transitions that can be implemented with both classical and quantum noise coincides if the amount of noise—quantified by the dimension  $m$ —is unbounded. In this case we have

$$\rho \xrightarrow{\infty} \rho' \Leftrightarrow \rho \xrightarrow{Q} \rho' \Leftrightarrow \rho \succ \rho', \quad (3)$$

where we use the symbol “ $\succ$ ” to indicate that  $\rho$  majorizes  $\rho'$  [11]. The set of transitions  $\rho \xrightarrow{\infty} \rho'$  have been extensively studied as noisy operations [1], where the noise is treated as a free resource and the main concern is to study the possible transitions with unbounded  $m$ . In contrast, here we are concerned with treating noise as a valuable resource and focus on the following question: What is the minimal amount of noise—quantified by  $m$ —that serves to implement any possible transition between pairs of  $d$ -dimensional quantum states fulfilling  $\rho \succ \rho'$ ? We denote these minimal values of  $d$  for the classical and quantum case by  $m_C^*(d)$  and  $m_Q^*(d)$ , respectively.

At first glance, one might suspect that  $m_C^*(d) = m_Q^*(d)$ , with quantum noise offering no advantage over its classical counterpart. That intuition comes from the fact that, although one writes a full quantum description in Eq. (2), the state of  $R$ , given by  $\mathbb{I}_m$ , is nevertheless a quasiclassical state. Hence, it seems reasonable that it could be recast as a classical variable, similarly as in Eq. (1). However, treating the noise as a quantum state allows one to access its quantum degrees of freedom, for example, to create entanglement between the  $S$  and  $R$ . In other words, one could in principle use quantum correlations to make a more efficient use of the noise yielding  $m_C^*(d) > m_Q^*(d)$ .

One of the main results of this work is to show that there is indeed a gap between the classical and quantum case. We find that  $m_C^*(d) = d > \lceil d^{1/2} \rceil = m_Q^*(d)$ , and more importantly, we construct protocols that saturate those bounds. In this way, we provide protocols that use the noise optimally for a large variety of tasks. These protocols also have a number of useful properties, such as allowing one to reuse the noise or being robust under different classes of imperfections. In the subsequent section, we present the key lemma to construct such optimal protocols and then turn to discuss applications and properties in Sec. IV.

## III. AN OPTIMAL DEPHASING MAP

For any state transition  $\rho \rightarrow \rho'$  that is possible under either quantum or classical noisy processes, there exists a corresponding map  $\mathcal{E}(\rho) = \rho'$  such that

$$\mathcal{E}(\cdot) = \mathcal{U}' \circ \pi_A \circ \mathcal{U}(\cdot). \quad (4)$$

Here,  $\mathcal{U}', \mathcal{U}$  are unitary channels that depend on  $\rho$  and  $\rho'$ . The map  $\pi_A$  is the dephasing map in a fixed orthonormal basis  $A = \{|i\rangle\}_{i=1}^d$ , defined as

$$\langle i | \pi_A(\rho) | j \rangle = \langle i | \rho | j \rangle \delta_{i,j}, \quad (5)$$

with  $\delta_{i,j}$  being the Kronecker delta. This follows from the Schur-Horn theorem [12] together with Eq. (3) and was used to bound the required randomness for noisy processes already in Ref. [13]. Since the unitary channels  $\mathcal{U}', \mathcal{U}$  do not require the use of any SOR by definition, we see from



Eq. (4) that noise is required only for the implementation of the dephasing map  $\pi_A$ . In turn, Eq. (4) implies that whether  $\mathcal{E}$  represents a quantum noisy process or a classical one depends only on the particular implementation of this dephasing map: Any construction of  $\pi_A$  in the form of Eq. (2) with  $m$ -dimensional SOR implies also that  $\mathcal{E}$  is a map  $\mathcal{E}_Q^m$ , while any construction of it in the form of Eq. (1) implies that  $\mathcal{E}$  is of the form  $\mathcal{E}_C^m$ .

Understanding the amount of randomness required to implement the dephasing map therefore is key to understanding the amount of randomness required to implement any noisy process. The following lemma provides a protocol implementing a dephasing map in any basis, using an explicit model of noise and requiring a SOR of dimension  $m = \lceil d^{1/2} \rceil$ .

*Lemma 1 (Catalytic quantum dephasing).*—For any integer  $d$  and basis  $A$  there exists a unitary  $U$ , so that

$$\mathrm{tr}_R[U(\cdot \otimes \mathbb{I}_{\lceil d^{1/2} \rceil})U^\dagger] = \pi_A(\cdot), \quad (6)$$

$$\mathrm{tr}_S[U(\rho \otimes \mathbb{I}_{\lceil d^{1/2} \rceil})U^\dagger] = \mathbb{I}_{\lceil d^{1/2} \rceil} \quad \forall \rho. \quad (7)$$

*Proof.*—Assume first that  $\sqrt{d} = m \in \mathbb{N}$ . Now, let  $\{U_i\}$  be a unitary operator basis for  $\mathcal{B}(\mathcal{H}_R)$ , that is, a collection of  $m^2 = d$  unitary operators  $U_i \in \mathcal{B}(\mathcal{H}_R)$ , such that

$$\frac{1}{m} \mathrm{tr}(U_i U_j^\dagger) = \delta_{i,j} \quad (8)$$

for all  $i, j$ . Such a basis exists for every  $m$  [14,15]. We now define the unitary,

$$U = \sum_{i=1}^d |i\rangle\langle i| \otimes U_i, \quad (9)$$

where the  $\{|i\rangle\}$  are elements of the basis  $A$  in which we intend to pinch. Then, for any density matrix  $\rho$  on  $\mathcal{H}_S$ ,

$$\mathrm{tr}_R[U(\rho \otimes \mathbb{I}_m)U^\dagger] = \sum_{i,j} |i\rangle\langle i| \rho |j\rangle\langle j| \frac{1}{m} \mathrm{tr}(U_i U_j^\dagger) \quad (10)$$

$$= \sum_{i,j} |i\rangle\langle i| \rho |j\rangle\langle j| \delta_{i,j} = \pi_A(\rho). \quad (11)$$

Lastly, note that Eq. (7) follows simply by

$$\mathrm{tr}_S[U(\rho \otimes \mathbb{I}_m)U^\dagger] = \sum_i \langle i|\rho|i\rangle U_i \mathbb{I}_m U_i^\dagger = \mathbb{I}_m. \quad (12)$$

In the case where  $\sqrt{d}$  is not an integer, we can use the same construction with a source of randomness of dimension  $m = \lceil d^{1/2} \rceil$  by simply not exhausting all possible  $m^2$  possible unitaries  $U_i$  on  $R$ . ■

The protocol of Lemma 1 is optimal, in the sense that it is impossible to implement the dephasing map with

$m < \lceil d^{1/2} \rceil$ . This can be seen by noting that for any basis  $A$  one can always choose an initial pure state  $\rho$  so that  $\pi_A(\rho) = \mathbb{I}_d$ . Using the preservation of the von Neumann entropy under unitaries and the Lieb-Araki triangle inequality, one finds that  $m \geq \sqrt{d}$  (see Appendix A). This implementation of the dephasing map compares with the best value known to date of  $m = d$ , proven in Ref. [13], whose implementation can in fact be shown to correspond to a classical noisy operation of the form Eq. (1), as we see later.

### A. Catalyticcity

Equation (7) states that the dephasing operation defined in Lemma 1 leaves the state of  $R$  invariant, or in other words, that the noise is catalytic [10,16–18]. This property has numerous useful applications. For instance, an immediate corollary of the lemma is that one can locally dephase an arbitrarily large number of uncorrelated systems, each of them of dimension at most  $d$ , by using a single noise system  $R$  of dimension  $\lceil d^{1/2} \rceil$ . More formally, we have that for any set of states  $\{\rho^i\}_{i=1}^N$  there exists a unitary  $U$  so that

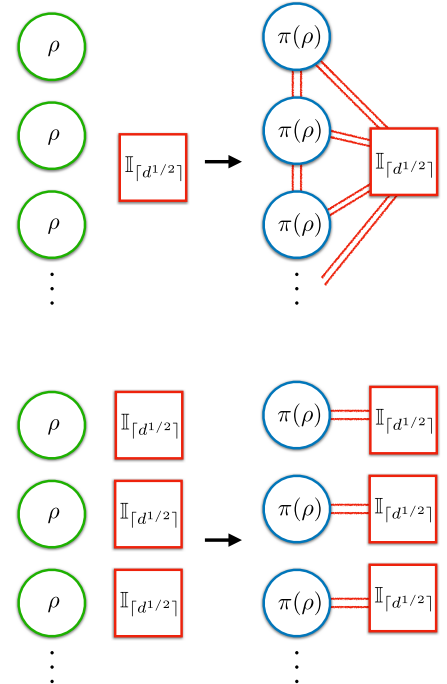


FIG. 1. Two possible ways of dephasing and the resulting correlation structure. Top: A sequence of systems in state  $\rho$  is dephased using a single state of randomness, with correlations being established between all systems involved. The local marginals of the resulting global state Eq. (13) are the dephased initial states. Bottom: In order to avoid correlations between the systems, one can instead use additional and unused randomness.

$$\mathrm{tr}_R[U(\rho_{S_1}^i \otimes \cdots \otimes \rho_{S_N}^i \otimes \mathbb{I}_{\lceil d^{1/2} \rceil})U^\dagger] = \rho_{S_1, \dots, S_N}^i, \quad (13)$$

where  $\rho_{S_i}^i = \pi_{A_i}(\rho_{S_i}^i)$ . This follows by simply iterating the unitaries of Lemma 1 with all the subsystems and reusing the noise, as illustrated in the top of Fig. 1. In contrast, if the noise would not have the property of being catalytic, then it would be necessary to employ a new mixed state for each of the subsystems, in which case an amount of randomness proportional to  $N$  would be required (bottom of Fig. 1). It is important to note, however, that reusing the randomness comes at the cost of correlating the subsystems amongst each other. Hence, if a protocol requires for the individual systems to remain uncorrelated, one still has to resort to a scheme whose required randomness scales linearly with the number of subsystems.

As sketched already, dephasing can be related to many processes that require noise, both in engineered as well as in equilibrating natural quantum processes. In the remainder of this work, we discuss and present applications of Lemma 1 to these processes.

#### IV. APPLICATIONS

##### A. Minimal noise for state transitions

As a first application, we prove the tight bounds for noisy operations presented in Sec. II. Formally, given a Hilbert space  $\mathcal{H}_S$  with  $\dim(\mathcal{H}_S) = d$ , we define the minimal noise for the classical and quantum case as

$$m_C^*(d) := \arg \min_m \rho \xrightarrow{m} \rho' \quad \forall \rho, \rho' \in \mathcal{B}(\mathcal{H}_S) | \rho \succ \rho', \quad (14)$$

$$m_Q^*(d) := \arg \min_m \rho \xrightarrow{m} \rho' \quad \forall \rho, \rho' \in \mathcal{B}(\mathcal{H}_S) | \rho \succ \rho'. \quad (15)$$

In the following lemma we find the values of the above quantities, thus providing the smallest SOR that suffices to perform any transition between two states  $\rho \succ \rho'$ . Note, however, that it is possible for particular transitions to require even less randomness or none at all.

*Lemma 2 (Optimal source of randomness for state transitions).*—Any state transition of a  $d$ -dimensional system that is possible under noisy processes, in the sense of Eqs. (14) and (15), can be implemented using an amount of classical and quantum noise given by

$$m_C^*(d) = d, \quad (16)$$

$$m_Q^*(d) = \lceil d^{1/2} \rceil. \quad (17)$$

*Proof.*—Here, we prove only that the above values are sufficient. For the corresponding necessary conditions (and  $\epsilon$ -approximate versions of the above), see Appendix A. Equation (17) follows from combining Eq. (4) with the dephasing construction in Lemma 1. To see Eq. (16), consider the unitary

$$V = \sum_{i=1}^d |i\rangle\langle i|_S \otimes X_R^i, \quad (18)$$

where  $X$  is the generalized Pauli matrix defined as

$$X|i\rangle = |(i+1) \bmod d\rangle. \quad (19)$$

As shown in Ref. [13], this unitary implements the dephasing map,

$$\mathrm{tr}_R[V(\rho \otimes \mathbb{I}_d)V^\dagger] = \frac{1}{d} \sum_{i,j} \langle i|\rho|j\rangle |i\rangle\langle j| \mathrm{tr}(X^{i-j}) = \pi_A(\rho). \quad (20)$$

$V$  is the local Fourier transform of a unitary leading to a channel of the form Eq. (1): there exists a unitary  $F$  and a basis  $\{|\tilde{j}\rangle = F^\dagger|j\rangle\}$  such that

$$\tilde{V} := (\mathbb{1} \otimes F)V(\mathbb{1} \otimes F^\dagger) = \sum_{j=1}^d Z^j \otimes |\tilde{j}\rangle\langle \tilde{j}|. \quad (21)$$

Here,

$$Z = \sum_j \omega_d^j |j\rangle\langle j| \quad (22)$$

is the generalized Pauli matrix conjugate to  $X$  and  $\omega_d$  the  $d$ th root of unity. Since the maximally mixed state is unitarily invariant,  $\tilde{V}$  implements the dephasing map, and its action on the system  $S$  can be represented as

$$\rho \mapsto \mathrm{tr}_R[\tilde{V}(\rho \otimes \mathbb{I}_d)\tilde{V}^\dagger] = \frac{1}{d} \sum_{j=1}^d Z^j \rho Z^{-j}. \quad (23)$$

Thus the dephasing map can be implemented with a classical SOR of dimension  $d$ . ■

This lemma proves a conjecture in Ref. [13], where the possibility of strengthening their bound  $m_Q^*(d) = d$  to the present one was already raised.

In complete analogy to the discussion in Sec. III A and Fig. 1, we can also use the catalytic properties of the source of randomness to implement state transitions locally from an initially uncorrelated state and using a fixed-size source of randomness. More concretely, let  $\{\rho^i\}_{i=1}^N$  and  $\{\sigma^i\}_{i=1}^N$  be  $d$ -dimensional quantum states such that  $\rho^i \succ \sigma^i$  for all  $i = 1, \dots, N$ . Then there exists a unitary  $U$  such that

$$\mathrm{tr}_R[U(\rho_{S_1}^1 \otimes \cdots \otimes \rho_{S_N}^N \otimes \mathbb{I}_{\lceil d^{1/2} \rceil})U^\dagger] = \rho_{S_1, \dots, S_N}^i, \quad (24)$$

with  $\rho_{S_i}^i = \sigma^i$ . To see this, we recall from the discussion in Sec. III A that the transition  $\rho^i \rightarrow \sigma^i$  can be implemented composing unitary channels and dephasing maps. Hence,  $\mathcal{E}(\rho_{S_1}^1 \otimes \cdots \otimes \rho_{S_1}^1) = \sigma_{S_1}^1 \otimes \cdots \otimes \sigma_{S_1}^1$ , with

$$\mathcal{E} = \bigotimes_{i=1}^N \mathcal{U}'_{S_i} \circ \bigotimes_{i=1}^N \pi_{A_i} \circ \bigotimes_{i=1}^N \mathcal{U}_{S_i}. \quad (25)$$

Now, using Eq. (13) we see that it is possible to dephase locally—that is, perform locally the same transition as the one implemented by the second map on the rhs of Eq. (25)—using a single source of randomness of dimension  $\lceil d^{1/2} \rceil$ , at the cost of creating correlations between the subsystems. Hence, composing the local unitaries with the local dephasing of Eq. (13), we obtain a map that locally implements the same transition as  $\mathcal{E}$ , as captured by Eq. (24).

### B. Smallest possible decohering environment and measurement device

A further application of our results is to the physical mechanism of decoherence and implementing a measurement in quantum mechanics, which can indeed be seen as a special case of a noisy operation, since it requires randomness. Both applications follow from the fact that a quantum source of randomness can be seen as half of a maximally entangled system.

It is useful to first discuss decoherence. To do so, we make use of the fact that the usual decoherence mechanism is, in a sense, simply a purified version of the system-environment interactions that are toy modeled by noisy operations. Let  $|\psi\rangle \in \mathcal{H}_S$  be an initial state vector of a  $d$ -dimensional system and  $|\phi\rangle$  be the initial state vector of the environment. According to the decoherence mechanism, the unitary joint evolution of system and bath is generated by a Hamiltonian whose interaction term picks out, or einselects, a preferred basis in which it decoheres the system [4]. We are now interested in the smallest possible size of the environment that achieves this. Let us label the system basis that is einselected by  $A = \{|i\rangle\}$  and assume that  $|\phi\rangle$  is a maximally entangled  $d$ -dimensional and bipartite state vector over systems  $E_1$  and  $E_2$ . We then define the unitary

$$U = U_{SE_1} \otimes \mathbb{1}_{E_2}, \quad (26)$$

where  $U_{SE_1}$  is the unitary defined in Eq. (9) that acts on systems  $S$  and  $E_1$ . As is clear from the above, this unitary will have the effect that

$$\text{tr}_E[U|\psi\rangle\langle\psi| \otimes |\phi\rangle\langle\phi|U^\dagger] = \pi_A(|\psi\rangle\langle\psi|), \quad (27)$$

meaning that even in this purified picture only an environment of the size of the system is required to produce decoherence.

Let us now turn to the smallest possible measurement device. For simplicity, we consider only projective measurement schemes: Suppose we are given a system in some initial state vector  $|\psi\rangle$  and some set of projective measurement operators  $\{M_i = |i\rangle\langle i|, i \in \{1, \dots, d\}\}$ . Then a

measurement process consists of the following steps. A bipartite measurement device, initially in state vector  $|\phi\rangle$ , consisting of a  $d$ -dimensional pointer system  $P$  and a remainder  $R$ , whose dimension we are interested in bounding, and a unitary  $W$  with the effect that

$$\text{Tr}_R[W|\psi\rangle\langle\psi| \otimes |\phi\rangle\langle\phi|W^\dagger] = \sum_i p_i |i, P_i\rangle\langle i, P_i|, \quad (28)$$

where  $p_i = \text{tr}(M_i|\psi\rangle\langle\psi|)$  and  $\{|P_i\rangle\}$  form an orthonormal basis for the pointer system. Using the above results, we can easily construct a measurement process as follows. Let the initial state vector of the measurement device be  $|\phi\rangle = |0\rangle_P \otimes |\phi^+\rangle_R$ , where  $|\phi^+\rangle$  is a bipartite,  $d$ -dimensional, maximally entangled state vector. Further, let  $\{V_i\}$  be unitaries defined by the action

$$V_i|i, 0\rangle = |i, P_i\rangle. \quad (29)$$

Finally, define the unitary

$$W = \sum_i |i\rangle\langle i| \otimes V_i \otimes (U_i)_{R_1} \otimes \mathbb{1}_{R_2}, \quad (30)$$

where the unitaries  $U_i$  form an operator basis as before. Then, it is easy to verify that  $|\phi\rangle$  and  $W$  together satisfy Eq. (28). This shows that in principle one requires a measurement device (including the pointer variable) whose size is only twice that of the system to be measured to implement a projective measurement as a physical process. Using entropic arguments one can again show that this is also the smallest possible measurement device. Note that the register  $R$  is exclusively used as a source of randomness in this protocol. Thus, if we are willing to give up the assumption that the initial state of the measurement device is pure, then it suffices to keep only part  $R_1$  in a maximally mixed state. Clearly, these results can also be read as providing the minimal dimension of an environment that equilibrates a quantum system of dimension  $d$  [2,3].

### C. Universal dephasing machine

In Sec. III, we show that with the aid of a noise system  $R$  in state  $\mathbb{I}_{\lceil d^{1/2} \rceil}$  it is possible to perform a protocol  $U$  which has the effect of implementing the dephasing map  $\pi_A$  on the system  $S$ . We now investigate which map is induced on  $S$  if the same unitary is applied with a system  $R$  in a state  $\sigma$  different from  $\mathbb{I}_{\lceil d^{1/2} \rceil}$ . We show that  $U$  brings the system closer to  $\pi_A(\rho)$  for any initial states  $\rho$  and  $\sigma$ . Also, we find that iterating the same protocol  $U$  with a sufficiently large sequence of imperfect noise states of  $R$  brings the system  $S$  exponentially close (in the number of iterations) to its dephased state. In this sense,  $U$  acts as a universal dephasing machine (Figs. 2 and 3): an iterated use of the same protocol  $U$  dephases the state of  $S$  for large families of states on  $R$  acting as a SOR. Hence, one can

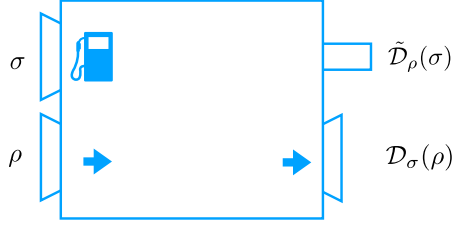


FIG. 2. Single instance of “universal dephasing machine.” We interpret the process  $\rho \otimes \sigma \rightarrow U(\rho \otimes \sigma)U^\dagger$  as a dephasing machine that takes the state  $\sigma$  as fuel and transfers the input state  $\rho$  into the output state  $\mathcal{D}_\sigma(\rho)$  and “waste”  $\tilde{\mathcal{D}}_\rho(\sigma)$ .

implement this protocol universally as a “black box,” without having to know the actual state of  $R$ .

### 1. Imperfect noise and convergence to the dephased state

Let  $\mathcal{D}_\sigma(\cdot)$  denote the map

$$\mathcal{D}_\sigma(\cdot) := \text{tr}_R[U(\cdot \otimes \sigma)U^\dagger], \quad (31)$$

where  $U$  is the unitary of Lemma 1. In Appendix B, we show that, for any  $\rho$  and  $\sigma$ ,

$$\mathcal{D}_\sigma(\pi(\rho)) = \pi(\mathcal{D}_\sigma(\rho)) = \pi(\rho), \quad (32)$$

$$\|\mathcal{D}_\sigma(\rho) - \pi(\rho)\|_1 \leq \|\sigma - \mathbb{I}_{[d^{1/2}]}\|_1, \quad (33)$$

where we have dropped the subscript  $A$ . These properties imply that, independently of the actual state  $\sigma$ , the system  $S$  is brought closer to the dephased state  $\pi(\rho)$  while keeping its diagonal invariant. This follows from the data-processing inequality [7]

$$\|\mathcal{D}_\sigma(\rho) - \pi(\rho)\|_1 = \|\mathcal{D}_\sigma(\rho) - \mathcal{D}_\sigma(\pi(\rho))\|_1 \leq \|\rho - \pi(\rho)\|_1.$$

Using those properties, one can show that by repeating the process sequentially (see Fig. 2, top) the system is eventually dephased for large classes of states  $\sigma$ . In fact, one can show that (see again Appendix B)

$$\|\mathcal{D}_\sigma^n(\rho) - \pi(\rho)\|_1 \leq \|\sigma - \mathbb{I}_{[d^{1/2}]}\|_1^n, \quad (34)$$

where  $\mathcal{D}_\sigma^n(\rho)$  denotes the repeated application of  $\mathcal{D}_\sigma$ . This means that, given  $\sigma$  such that  $\|\sigma - \mathbb{I}_{[d^{1/2}]}\|_1 < 1$ , the dephased state is approached exponentially fast. Note that another corollary of the above properties is that the map  $\mathcal{D}_\sigma$  can only increase the von Neumann entropy of its input, which is formally proven in Appendix B 1.

### 2. Reusing the randomness

In the case of  $R$  being in the state  $\mathbb{I}_{[d^{1/2}]}$ , we show in Sec. III A that it remains unchanged and, thus, the noise is reusable. A natural question is then what happens to the state of  $R$  when it is in an arbitrary state  $\sigma$ . Let  $\tilde{\mathcal{D}}_\rho$  denote the map

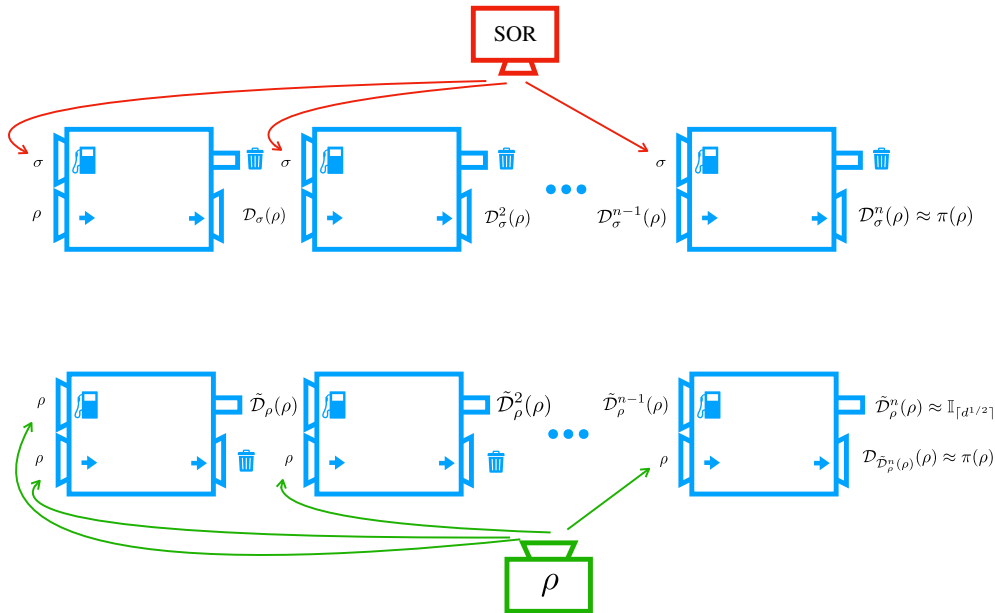


FIG. 3. Top: Repeated application on single input state approximates dephasing map. Bottom: Producing the dephased state when there is no SOR. If  $\|\rho - \mathbb{I}_d\|_1 < 1$ , then the necessary amount of randomness for dephasing can be distilled by repeated application of the universal dephasing machine.

$$\tilde{\mathcal{D}}_\rho(\cdot) := \text{tr}_R[U(\rho \otimes \cdot)U^\dagger]. \quad (35)$$

It follows simply from Eq. (12) that  $\tilde{\mathcal{D}}_\rho$  is just a mixture of unitaries, hence bringing  $R$  closer to the maximally mixed state. Indeed, following arguments analogous to the ones of Sec. IV C 1 (see Appendix B), one can show that there exist choices for the unitary operator basis of Lemma 1 so that the final state of  $R$  fulfills

$$\|\tilde{\mathcal{D}}_\rho(\sigma) - \mathbb{I}_{\lceil d^{1/2} \rceil}\|_1 \leq \|\rho - \mathbb{I}_d\|_1, \quad (36)$$

and analogously it converges as

$$\|\tilde{\mathcal{D}}_\rho^n(\sigma) - \mathbb{I}_{\lceil d^{1/2} \rceil}\|_1 \leq \|\rho - \mathbb{I}_d\|_1^n. \quad (37)$$

Altogether we conclude not only that the noise can be reused, but furthermore, that it improves its quality converging exponentially fast to a state of perfect noise, provided that the initial state  $\rho$  is mixed enough to start with (as given by the condition  $\|\rho - \mathbb{I}_d\|_1 < 1$ ). The fact that the noise system is brought closer to the maximally mixed state allows one to implement a distillation protocol such as the one depicted in Fig. 3 (bottom). There, one has a single source providing copies of a given initial state  $\rho$ . One aims at dephasing each subsystem locally, similarly to what is done with a perfect noise system in Eq. (13). Here, one can take one copy  $\rho$  playing the role of  $R$  for some iterations until it is brought close enough to the maximally mixed state, which will happen exponentially quickly, given Eq. (37). Then, using Eq. (34), one can ensure that all the new copies of  $\rho$  can be locally dephased.

### 3. Time control for the dephasing machine and recurrence

Thus far we have left unspecified how the dephasing of the machine would physically be implemented. One concern here may be that the dephasing properties heavily rely on very precise time control of the evolution under the associated Hamiltonian  $H = i \log(U)$ . However, the numerical simulations depicted in Fig. 4 strongly indicate that, as the system dimension becomes large,  $H$  produces an evolution that is close to  $\mathcal{D}_\sigma(\cdot)$  for a time span that scales exponentially with the size of  $S$ . Indeed, for prime power dimensions and the case  $\sigma = \mathbb{I}_{\lceil d^{1/2} \rceil}$ , we find analytically that integer iterations of the application of the dephasing unitary always yield the exact dephasing map, up to a recurrence point, at which the original state is returned. See Appendix C for details. The numerical simulations above complement this and suggest that this recurrence property holds not only for integer iterations of the application of the dephasing unitary, but also for intermediate times.

We hence expect that in the limit of very large dimensions, this equilibrating behavior [2,3] becomes arbitrarily

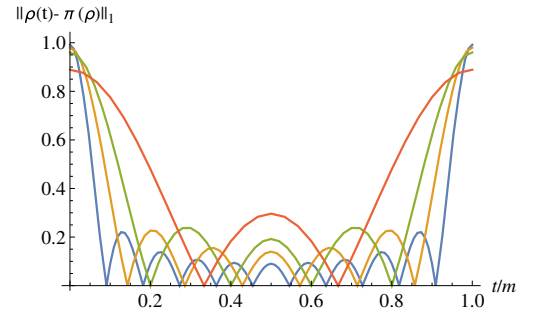


FIG. 4. Numerical simulations of the dephasing map that is induced by the noisy operation Eq. (9) for continuous time and system dimensions  $d = m^2 = 9, 25, 49, 121$  (red, green, yellow, blue lines). Shown is the trace-norm distance between the time-evolved state  $\rho(t)$  and the pinched state  $\pi(\rho)$  as a function of rescaled time  $t/m$ . The initial state is a maximally coherent state  $(1/\sqrt{d}) \sum_i |i\rangle$ . The graph shows that, while for integer times (with respect to the bath dimension) the dephasing is always exact, for noninteger times the deviation from exact dephasing becomes small with increasing dimension. The numerically obtained deviation at  $t/m = 0.5$  seems compatible with a scaling as  $1/m = 1/\sqrt{d}$ , but we leave open to derive the exact scaling behavior.

good and the state  $\rho(t)$  remains close to the equilibrium state  $\pi(\rho)$  for a time exponential in the system size. This means that the universal dephasing machine can be made robust in time, in the sense that it does not require exact control over the timing and the dephasing is maintained for long timescales.

### D. Entanglement-assisted private quantum channel

In this section, we apply our results to the construction of a cryptographic protocol known as a private quantum channel (PQC). In a PQC setting, two parties, Alice and Bob, would like to communicate quantum data privately, that is, without an eavesdropper being able to intercept and retrieve the data. To achieve this they share a secret key. We now first briefly explain PQCs using classical secret keys and then provide a construction where the classical key  $k$  is substituted for a “quantum key” in the form of a minimal number of entangled bits. In the following, we denote by  $\mathcal{S}(\mathcal{H})$  the set of normalized quantum states on the Hilbert space  $\mathcal{H}$ . Formally, in the classical setting, a  $(\delta, \epsilon)$  PQC is a set of pairs of encoding and decoding completely positive trace-preserving (CPTP) maps  $\mathcal{X}_k: \mathcal{S}(\mathcal{H}_A) \rightarrow \mathcal{S}(\mathcal{H}_{A'})$  and  $\mathcal{Y}_k: \mathcal{S}(\mathcal{H}_{A'}) \rightarrow \mathcal{S}(\mathcal{H}_A)$  that can be locally implemented by the sending and receiving parties, respectively, where  $k$  denotes the secret key that is shared by Alice and Bob. We think of the key  $k$  as a random variable and assume that the key  $k$  occurs with probability  $p(k)$ . These channels then have to fulfill the following conditions [19]. Firstly, there exists a fixed element  $\tau \in \mathcal{S}(\mathcal{H}_{A'})$ , such that



$$\sup_{\rho_{A,B} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)} \left\| \left( \sum_k p_k \mathcal{X}_k \otimes \text{id} \right) (\rho_{A,B}) - \tau \otimes \rho_B \right\|_1 \leq \epsilon, \quad (38)$$

where  $\rho_{A,B}$  is any extension of the input state  $\rho_A$  to a larger Hilbert space and  $\rho_B = \text{tr}_A(\rho_{A,B})$ . And secondly,

$$\sup_{\rho \in \mathcal{S}(\mathcal{H}_A)} \left\| \sum_k p_k \mathcal{Y}_k \circ \mathcal{X}_k(\rho) - \rho \right\|_1 \leq \delta. \quad (39)$$

Equation (38) warrants (approximate) security from eavesdropping, while Eq. (39) warrants the channel's (approximate) reliability. The reason that the security is defined over all possible extensions is that the eavesdropper may initially be entangled with part of the unencrypted message. Finally, a (0,0) PQC is called an *ideal* PQC.

PQCs have been well studied for the case in which Alice and Bob share a classical key [8,9,19–22]. In this case, and if  $\mathcal{X}_k$  is unitary, the encoding corresponds to a classical noisy process and a key of length at least  $[2 - O(\epsilon)]n$  is necessary for the  $\epsilon$ -secure transmission of  $n$  qubits [8,9,19,23].

Here, in contrast, we consider a setting in which Alice and Bob share a quantum key in the form of entangled quantum states. We use our dephasing map to construct an ideal private quantum channel that requires  $n$  shared ebits of entanglement to transmit  $n$  qubits of quantum data. As with the dephasing map, this value can again be shown to be optimal, in the sense that no implementation of an ideal PQC as a noisy operation can require fewer ebits (a result that extends to approximately ideal PQCs). It improves on the only other discussion of PQCs that uses entanglement known to the authors, in Ref. [25]. There, an ideal PQC is constructed that applies techniques from classical PQCs and hence achieves only ‘‘classical’’ efficiency by requiring  $2n$  ebits for  $n$  transmitted qubits.

The idea behind our construction is straightforward (see Fig. 5). Given an  $n$ -qubit system  $S$ , let  $U_I$  and  $U_J$  denote the dephasing unitaries Eq. (9) whose projective part corresponds to the two orthonormal bases  $I = \{|i\rangle\}_{i=1}^d$  and  $J = \{|j\rangle\}_{j=1}^d$  for  $\mathcal{H}_S$ . If Alice and Bob share  $n$  ebits, and assuming for convenience that  $n$  is even, Alice can split the ebits into two halves, which we call  $E_1$  and  $E_2$ . She then applies  $U_I$  to  $S$  and her local share of  $E_1$ , followed by applying  $U_J$  to  $S$  and her half of  $E_2$ . It is easy to check that if  $I$  and  $J$  are mutually unbiased, that is, if

$$|\langle i|j\rangle|^2 = \frac{1}{d}, \quad \forall i, j, \quad (40)$$

then this results in the completely depolarizing channel. That is, the map

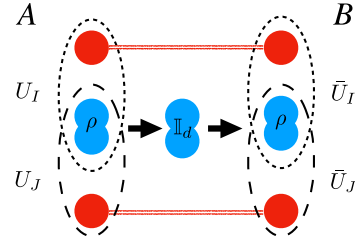


FIG. 5. Illustration of our quantum PQC for the case  $n = 2$ . To encode a 2-qubit state  $\rho$  (blue), Alice applies the dephasing unitaries  $U_I$  and  $U_J$  to the system and one half of an ebit (red) each, where  $I$  and  $J$  can be any mutually unbiased bases. This maps  $\rho$  into the maximally mixed state exactly, so that an eavesdropper cannot learn anything about  $\rho$  even if she was initially entangled with part of it. Bob, in order to decode, applies the conjugate of the two above unitaries and thereby retrieves the state exactly.

$$\mathcal{X}(\cdot) := \text{tr}_E[U_J U_I(\cdot \otimes |\phi^+\rangle\langle\phi^+|_{E_1} \otimes |\phi^+\rangle\langle\phi^+|_{E_2}) U_I^\dagger U_J^\dagger], \quad (41)$$

where  $|\phi^+\rangle$  represents an  $n/2$ -ebit state vector, has the property that

$$\mathcal{X}(\rho) = \mathbb{I}_d, \quad \forall \rho \in \mathcal{D}(\mathcal{H}_S). \quad (42)$$

This ensures perfect secrecy, since the completely depolarizing channel necessarily also removes all correlations to other systems [20]. Upon receipt of  $S$ , Bob can then apply the complex conjugate of the encoding unitaries to his share of the ebits to retrieve the original state. See Appendix D for the formal proofs.

This construction has a number of interesting features, some of which, however, are already present in the construction of Ref. [25]. For instance, it is catalytic in the sense that, at the end of the transmission process, in case no eavesdropper has interacted with the sent data, all of the entanglement is returned in its initial state and can be reused for future rounds of transmission. Moreover, the scheme allows for error correction, efficient authentication, and recycling of some of the entanglement in case eavesdropping has occurred. We refer the reader to Appendix D for a discussion of these properties.

## V. DEPHASING WITH QUANTUM EXPANDERS

The protocol presented in Lemma 1 allows one to dephase perfectly a  $d$ -dimensional system given a SOR of dimension of  $m = \lceil d^{1/2} \rceil$ . This very same protocol, when applied to an imperfect SOR of dimension  $m$  but not in the maximally mixed state, yields, as shown in Sec. IV C 1, a convergence to the dephased state when the protocol is iterated. In this section, we study a complementary protocol that provides astonishingly fast convergence when we have states of the SOR that are maximally

mixed, but of dimension significantly smaller than  $m$ . We find a protocol that yields an exponential convergence to the dephased state with the dimension of the SOR, measured in the 2-norm. This is remarkable, in that it shows that one can obtain an equilibration in 2-norm exponentially quickly in the ancillary dimension. This insight may be seen as being at odds with the intuition that an equilibrating environment should naturally have a large physical dimension. Our approach is based on a machinery of quantum expanders [26–28]. The key insight is that one can trade residual correlations still present in the system with the dimension required for the mixing environment. This feature demonstrates an intriguing feature of randomness.

*Theorem 3 (Dephasing with quantum expanders).*—For any  $d$ -dimensional state  $\rho$ ,  $d = e^2$  with  $d$  odd, and an integer  $k$ , there exists an  $8k$ -dimensional quantum system  $R$  and a unitary  $U \in (8dk)$ , such that

$$\|\text{tr}_R[U(\rho \otimes \mathbb{I}_{8k})U^\dagger] - \pi(\rho)\|_2 \leq \sqrt{2d^3} \left(\frac{5\sqrt{2}}{8}\right)^k. \quad (43)$$

The restriction to the dimension is done for pure conceptual simplicity. The argument for the proof, presented in Appendix E, follows from a construction of a classical random walk that acts on the vertices of an expander graph, a Margulis expander [29]. In the present construction, the vertices of the Margulis expander are seen as lines labeled by  $q = 1, \dots, d$  in a  $(d \times d)$ -dimensional quantum phase space of the  $d$ -dimensional quantum system. The central insight is that classical random walks on such lattices are reflected by random walks on Wigner functions defined on  $(d \times d)$ -dimensional phase spaces, which in turn give rise to random unitary channels on quantum states in  $d$  dimensions. The construction laid out in detail in the Appendix E builds upon and draws inspiration from the scheme of Ref. [27], but is in several important ways a new scheme, in particular, in that each line in phase space is treated separately. In this way, the strong mixing properties of the random walk of the Margulis expander graph are not used to show rapid mixing to a maximally mixed state, but in fact to a quantum state with vanishing off-diagonal elements.

## VI. SUMMARY AND CONCLUSIONS

We study the problem of implementing state transitions under noisy processes, that is, processes that require randomness. We solve this problem completely by providing optimal protocols for both the case of an implicit, classical model of randomness as well as an explicit, quantum model of randomness. The main building block behind these protocols is the construction of a protocol that performs a dephasing map on an arbitrary quantum state using a SOR of the smallest possible dimension, for both the quantum and classical case. We find that a quantum SOR is quadratically more efficient than its classical

counterpart due to quantum correlations, and hence show that an explicit model is strictly more powerful for any dimension  $d > 2$ .

Once the optimal protocols for dephasing were established, we studied applications such as state transitions in noisy operations, decoherence, and quantum measurements, providing optimal protocols for all of them. An interesting feature of our protocol is that the SOR is not altered during the protocol, meaning that it can be reused to implement further iterations of the above tasks.

We also extend our discussion to the case of imperfect noise and use our results to construct a universal dephasing machine that exhibits robustness both with respect to the noise that fuels it, as well as with respect to the control over timing when running it. Moreover, we use our dephasing as a primitive to construct a novel, ideal private quantum channel. Finally, by putting it into the context of expander graphs, we have seen how such an approximate dephasing is possible with an economical use of noise: Converging in 2-norm to the dephased state with an exponential scaling on the SOR’s dimension.

Besides the foundational interest of our construction, which makes precise the way in which the relationship between correlations and randomness in quantum mechanics differs from that in classical mechanics, we expect our dephasing protocol to improve bounds in noisy processes that we have not discussed here, to the extent that introduce a new primitive to constructions in quantum information. Given the pivotal status of randomness in protocols of quantum information processing and in notions of quantum thermodynamics, these results promise a significant number of further practical applications.

## ACKNOWLEDGMENTS

P. B. thanks Lluís Masanes, Markus Müller, Jon Richens, and Ingo Roth for interesting conversations, and especially Jonathan Oppenheim for suggesting cryptographic applications of the results. We acknowledge funding from the ERC (TAQ), the DFG (EI 519/14-1, CRC183, FOR 2724), the Templeton Foundation, and the Studienstiftung des Deutschen Volkes. This work has also received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No. 817482 (PASQUANS).

## APPENDIX A: LOWER BOUNDS ON DIMENSION OF SOURCE OF RANDOMNESS

In this appendix, we prove the lower bounds in Lemma 2. In fact, we prove them in an approximate setting to show that they are robust to small deviations from exact dephasing. To do so, we call a map  $\mathcal{E}_X^m$   $\epsilon$  dephasing if, for all operators  $\rho \in \mathcal{B}(\mathcal{H}_S)$  and some fixed basis  $A$ ,

$$\|\mathcal{E}_X^m(\rho) - \pi_A(\rho)\|_1 \leq \epsilon, \quad (A1)$$

where  $X \in \{C, Q\}$ . Let  $m_X^*(d, \epsilon)$  be the smallest value of  $m$  such that an  $\epsilon$ -dephasing map can be realized as a map of the form Eq. (1) for  $X = C$  and Eq. (2) for  $X = Q$ , respectively,  $\dim(\mathcal{H}_S) = d$ .

We begin with the classical bound. Consider the state vector

$$|A\rangle := \frac{1}{\sqrt{d}} \sum_i |i\rangle. \quad (\text{A2})$$

If it is dephased in the basis  $A = \{|i\rangle\}$ , it is mapped to the maximally mixed state. We are concerned with deriving the minimal value of  $m$  such that  $\mathcal{E}_C^m(|A\rangle\langle A|) = \mathbb{I}_d$ . For this, note that

$$\mathcal{E}_C^m(|A\rangle\langle A|) = \frac{1}{m} \sum_{j=1}^m U_j |A\rangle\langle A| U_j^\dagger. \quad (\text{A3})$$

Clearly, this state has at most rank  $m$ , since its support is spanned by  $m$  vectors. Moreover, it is easy to see that for any  $\epsilon$ -dephasing classical map  $\mathcal{E}_C^m$ ,

$$\text{rank } \mathcal{E}_C^m(|A\rangle\langle A|) \geq d \left(1 - \frac{\epsilon}{2}\right), \quad (\text{A4})$$

which implies

$$m_C^*(d, \epsilon) \geq m \geq \max \left\{ 2, d \left(1 - \frac{\epsilon}{2}\right) \right\}, \quad (\text{A5})$$

where we also use that any nontrivial source of randomness must be at least two dimensional.

To see Eq. (A4), consider any state  $\rho$  of rank  $k$ . Then,

$$\|\rho - \mathbb{I}_d\|_1 \geq \|\mathbb{I}_{k,d} - \mathbb{I}_d\|_1 = 2 \left(1 - \frac{k}{d}\right), \quad (\text{A6})$$

where  $\mathbb{I}_{k,d}$  is a  $d$ -dimensional state that is maximally mixed on a subspace of dimension  $k$  (and hence has rank  $k$ ). Using Eq. (A1) and rearranging then gives bound Eq. (A4).

Let us now turn to the quantum case, where we find

$$m_Q^*(d, \epsilon) \geq \max \{2, d^{(1-\epsilon)/2} \epsilon^{\epsilon/2}\}, \quad \forall \epsilon \leq \frac{1}{6e}. \quad (\text{A7})$$

First, note that for  $d \leq 4$ , our optimal construction already yields  $m = 2 = \lceil d^{1/2} \rceil$  and that any nontrivial source of randomness must have  $m \geq 2$ . In the following, we hence assume  $d \geq 5$ . Now consider again the initial state  $|A\rangle\langle A|$ . Then, for any  $\epsilon$ -dephasing map  $\mathcal{E}_Q^m$ , applying Fannes's inequality yields

$$S[\mathcal{E}_Q^m(|A\rangle\langle A|)] \geq \log d + \epsilon \log \left(\frac{\epsilon}{d}\right). \quad (\text{A8})$$

In the following, let  $\rho'_R$  denote the state on the  $m$ -dimensional source of randomness after the dephasing map has been applied. From our construction of the exact dephasing map, we know that  $m^*(d, \epsilon) \leq \lceil d^{1/2} \rceil$ . Hence, in the following we assume  $2 \leq m \leq \lceil d^{1/2} \rceil$ . Since  $\epsilon \leq 1/6e$  and

$$\log(\lceil d^{1/2} \rceil) - \log(d^{1/2}) \leq 1/2, \quad \forall d \geq 5, \quad (\text{A9})$$

it follows using Eq. (A8) that

$$S[\mathcal{E}_Q^m(|A\rangle\langle A|)] > \log(\lceil d^{1/2} \rceil) \geq S(\rho'_R). \quad (\text{A10})$$

We finally use the Lieb-Araki triangle inequality, which states that

$$S(\rho_{A,B}) \geq |S(\rho_A) - S(\rho_B)|, \quad (\text{A11})$$

for any bipartite state  $\rho_{A,B}$ . We can now use this to bound

$$\log m = S(|A\rangle\langle A|) + S(\mathbb{I}_m) = S(U|A\rangle\langle A| \otimes \mathbb{I}_m U^\dagger) \quad (\text{A12})$$

$$\geq |S[\mathcal{E}_Q^m(|A\rangle\langle A|)] - S(\rho'_R)| \quad (\text{A13})$$

$$= S[\mathcal{E}_Q^m(|A\rangle\langle A|)] - S(\rho'_R) \quad (\text{A14})$$

$$\geq \log(d) + \epsilon \log(\epsilon/d) - \log m. \quad (\text{A15})$$

Hence, we obtain

$$m \geq d^{(1-\epsilon)/2} \epsilon^{\epsilon/2}, \quad (\text{A16})$$

which finishes the proof.

## APPENDIX B: UNIVERSAL DEPHASING MACHINE

In this appendix, we provide further details on the results regarding the universal dephasing machine. For convenience, we drop the subscripts for the dephasing maps and the maximally mixed states.

### 1. Robustness with respect to imperfect noise

Let us first show the following lemma.

*Lemma 4 (General properties of  $\mathcal{D}_\sigma$ ).*—The family of channels  $\mathcal{D}_\sigma$  has the following properties.

- (1) Fixed points. All diagonal states are fixed points:

$$\mathcal{D}_\sigma(\pi(\rho)) = \pi(\rho), \quad \forall \sigma, \rho. \quad (\text{B1})$$

- (2) Invariant diagonal. The channels do not modify the diagonal of any state in the given basis:

$$\pi(\mathcal{D}_\sigma(\rho)) = \pi(\rho), \quad \forall \sigma, \rho. \quad (\text{B2})$$



(3) Continuity. The following continuity property holds:

$$\|\mathcal{D}_\sigma(\rho) - \pi(\rho)\|_1 \leq \|\sigma - \mathbb{I}\|_1. \quad (\text{B3})$$

*Proof.*—The first two properties follow from the definition of  $\mathcal{D}_\sigma$  in Eq. (31), since

$$\langle k | \text{tr}_B[U(\rho \otimes \sigma)U^\dagger] | k \rangle = \sum_{i,j} \langle k | i \rangle \langle i | \rho | j \rangle \langle j | k \rangle \text{tr}(U_i \sigma U_j^\dagger) \quad (\text{B4})$$

$$= \rho_{k,k} \text{tr}(U_k^\dagger U_k \sigma) = \rho_{k,k}. \quad (\text{B5})$$

The continuity property can be seen as

$$\begin{aligned} \|\mathcal{D}_\sigma(\rho) - \pi(\rho)\|_1 &= \|\text{tr}_B[U(\rho \otimes \sigma)U^\dagger] - \text{tr}_B[U(\rho \otimes \mathbb{I})U^\dagger]\|_1 \\ &\leq \|U(\rho \otimes (\sigma - \mathbb{I}))U^\dagger\|_1 \\ &= \|\rho \otimes (\sigma - \mathbb{I})\|_1 \\ &= \|\sigma - \mathbb{I}\|_1, \end{aligned} \quad (\text{B6})$$

where we have used the data-processing inequality and the unitary invariance of the norm. ■

In particular, the fixed-point property has the following corollaries.

*Corollary 5 (Contraction to dephased state).*—Let  $f(\rho, \rho')$  be any measure of distance between quantum states that fulfills the data-processing inequality, for example, any Renyi divergence or the trace distance [7]. Then,

$$f(\rho, \pi(\rho)) \geq f(\mathcal{D}_\sigma(\rho), \pi(\rho)), \quad \forall \sigma. \quad (\text{B7})$$

Choosing  $f(\rho, \sigma)$  as the quantum relative entropy  $S(\rho \| \sigma)$  and using that  $S(\rho \| \pi(\rho)) = S(\pi(\rho)) - S(\rho)$ , we then obtain the following corollary.

*Corollary 6 (Increasing entropy).*—The channels  $\mathcal{D}_\sigma$  can only increase the von Neumann entropy:

$$S(\rho) \leq S(\mathcal{D}_\sigma(\rho)), \quad \forall \sigma. \quad (\text{B8})$$

So far we have considered only single applications of the dephasing map. Let us now consider repeated applications. We thus want to investigate what happens if we have a stream of sources of randomness  $\sigma_i$  and sequentially use them to dephase the system. To this end, we can prove the following lemma.

*Lemma 7 (Iterated dephasing).*—Let  $\{\sigma_i\}_{i=1}^n$  be arbitrary quantum states of dimension  $\lceil d^{1/2} \rceil$ . Then we have

$$\|(\mathcal{D}_{\sigma_n} \circ \dots \circ \mathcal{D}_{\sigma_1})(\rho) - \pi(\rho)\|_1 \leq \prod_{i=1}^n \|\sigma_i - \mathbb{I}\|_1. \quad (\text{B9})$$

*Proof.*—We prove the case  $n = 2$ . The general result follows by iteration. First we use  $\pi(\rho) = \pi \circ \mathcal{D}_\sigma(\rho) = \mathcal{D}_\sigma \circ \pi(\rho)$  to write

$$\|(\mathcal{D}_{\sigma_2} \circ \mathcal{D}_{\sigma_1})(\rho) - \pi(\rho)\|_1 = \|(\mathcal{D}_{\sigma_2} - \pi) \circ (\mathcal{D}_{\sigma_1} - \pi)(\rho)\|_1.$$

We can then estimate this norm as

$$\|(\mathcal{D}_{\sigma_2} \circ \mathcal{D}_{\sigma_1})(\rho) - \pi(\rho)\|_1 \leq \|\mathcal{D}_{\sigma_1} - \pi\|_{1 \rightarrow 1} \|\mathcal{D}_{\sigma_2} - \pi\|_{1 \rightarrow 1}, \quad (\text{B10})$$

where  $\|\cdot\|_{1 \rightarrow 1}$  is the norm on superoperators induced by the 1-norm. From Lemma 4, we can estimate it as

$$\|\mathcal{D}_\sigma - \pi\|_{1 \rightarrow 1} = \max_\rho \|\mathcal{D}_\sigma(\rho) - \pi(\rho)\|_1 \leq \|\sigma - \mathbb{I}\|_1. \quad (\text{B11})$$

This step completes the proof. ■

We thus find that  $\rho$  converges exponentially quickly to the dephased state upon iterated application of  $\mathcal{D}_\sigma$  provided that  $\|\sigma_i - \mathbb{I}\|_1 \leq k < 1$  for some  $k$  and all  $\sigma_i$ .

## 2. Action on source of randomness

Let us now consider the action of the dephasing unitary on the source of randomness. Given some  $\rho$ , we are thus interested in the channel

$$\tilde{\mathcal{D}}_\rho(\sigma) = \text{tr}_S[U(\rho \otimes \sigma)U^\dagger]. \quad (\text{B12})$$

This channel is always unital; i.e., it fulfills  $\tilde{\mathcal{D}}_\rho(\mathbb{I}) = \mathbb{I}$  for any  $\rho$ . Thus,

$$\|\tilde{\mathcal{D}}_\rho(\sigma) - \mathbb{I}\|_1 \leq \|\sigma - \mathbb{I}\|_1. \quad (\text{B13})$$

Let us denote by  $\mathcal{R}$  the channel that maps any state into the maximally mixed state,  $\mathcal{R}(\sigma) = \mathbb{I}$ . Then we have  $\mathcal{R} = \tilde{\mathcal{D}}_\rho \circ \mathcal{R} = \mathcal{R} \circ \tilde{\mathcal{D}}_\rho$ . By the same arguments as in the previous section, we then obtain the following lemma.

*Lemma 8 (Iterated mixing).*—Let  $\{\rho_i\}_{i=1}^n$  be arbitrary quantum states of dimension  $d$ . Then we have

$$\|(\tilde{\mathcal{D}}_{\rho_n} \circ \dots \circ \tilde{\mathcal{D}}_{\rho_1})(\sigma) - \mathbb{I}\|_1 \leq \prod_{i=1}^n \|\rho_i - \mathbb{I}\|_1. \quad (\text{B14})$$

## APPENDIX C: RECURRENCE AND ROBUSTNESS IN TIME

In this appendix, we show that one can choose the operator basis  $\{U_i\}$  from Lemma 1 in such a way that the dephasing map exhibits recurrence properties. By recurrence we here mean that applying the dephasing unitary a certain number of times undoes the dephasing, while it keeps it dephased for intermediate times.

To this end, note that one particular realization of this operator basis is the following: Define the unitaries

$$U_{r,s} := \tau^{rs} X^r Z^s, \quad (\text{C1})$$

where  $X, Z$  are the generalized Pauli matrices defined in Eq. (19) and (22), respectively, and  $\tau = -e^{\pi i/m} = -\sqrt{\omega}$ .

In the following, expressions are to be taken modulo  $m$ , unless specified otherwise. The conjugation relation  $XZ = \omega^{-1}ZX$  then gives rise to the following properties in any dimension [30]:

$$U_{r,s}U_{u,v} = \omega^{us-vr}U_{u,v}U_{r,s} = \tau^{us-vr}U_{r+u,s+v}, \quad (\text{C2})$$

$$U_{r,s}^k = U_{kr,ks}, \quad (\text{C3})$$

$$U_{r,s}^\dagger = U_{-r,-s}, \quad (\text{C4})$$

$$\text{tr}(U_{r,s}) = m\delta_{r,0}\delta_{s,0}. \quad (\text{C5})$$

These imply, in particular, that  $\{U_{r,s}\}, r, s \in \{0, \dots, m-1\}$  form a unitary operator basis of  $\mathcal{B}(\mathcal{H})$ . Now, while it is clear that  $X^m = Z^m = \mathbb{I}$ , we can ask for the smallest  $k$  such that  $U_{r,s}^k = \mathbb{I}$  for all  $r, s$ . The above conjugation relations imply that if  $m$  is odd, then this value is given by  $m$ , while for even  $m$ , the answer is  $2m$ . For instance, in the case of  $m = 2$ , we have  $X^2 = Z^2 = \mathbb{I}$ , while  $(XZ)^2 = -\mathbb{I}$ . Moreover, we can ask for the dependence of the order of the unitaries  $U_i$ , by which we here mean the smallest  $k$  such that  $U_i^k = \mathbb{I}$ , i.e., the order of the corresponding element in the Weyl-Heisenberg group, on  $m$ . Here, one has that the order of all nontrivial  $U_i$  is  $d$ , if and only if  $d$  is an odd prime. This special property for odd primes will be of key importance to establish recurrence relations in the following. Define the map

$$\pi_m^k(\cdot) = \begin{cases} \text{id}(\cdot) & \text{if } k \bmod m = 0 \\ \pi_A(\cdot) & \text{otherwise,} \end{cases} \quad (\text{C6})$$

where  $A$  denotes the orthonormal basis in which the pinching acts, as in the main text. We then have the following lemma.

*Lemma 9 (Recurrence for odd prime dimension).*—Let  $\dim \mathcal{H}_S = m^2$ ,  $\dim \mathcal{H}_R = m$ , where  $m$  is an odd prime. There exists a unitary  $V$  acting on  $\mathcal{H}_S \otimes \mathcal{H}_R$  such that

$$\text{tr}_B[V^k(\rho \otimes \mathbb{I}_m)(V^\dagger)^k] = \pi_m^k(\rho). \quad (\text{C7})$$

*Proof.*—Let  $A = \{|r, s\rangle\}_{r,s=1}^m$  be the orthonormal basis of  $\mathcal{H}_S$  in which we want to pinch the state  $\rho$ . Define

$$V = \sum_{r,s} |r, s\rangle \langle r, s|_S \otimes (U_{r,s})_R, \quad (\text{C8})$$

where the basis with respect to which the operators Eq. (C1) are defined can be chosen arbitrarily. Then, from the properties of these operators, we have

$$\begin{aligned} \text{tr}_R[V^k(\rho \otimes \mathbb{I}/d)(V^\dagger)^k] &= \sum_{r,s,u,v} |r, s\rangle \langle r, s| \rho |u, v\rangle \langle u, v| \frac{1}{m} \text{tr}(U_{kr,ks}U_{-ku,-kv}) \\ &= \sum_{r,s,u,v} |r, s\rangle \langle r, s| \rho |u, v\rangle \langle u, v| \frac{1}{m} \tau^{k^2(us-rv)} \text{tr}(U_{r-u,s-v}^k) \end{aligned} \quad (\text{C9})$$

$$= \sum_{r,s,u,v} |r, s\rangle \langle r, s| \rho |u, v\rangle \langle u, v| \theta_m(k, r, u, s, v) \quad (\text{C10})$$

$$= \sum_{r,s,u,v} |r, s\rangle \langle r, s| \rho |u, v\rangle \langle u, v| \theta_m(k, r, u, s, v) \quad (\text{C11})$$

$$= \pi_m^k(\rho), \quad (\text{C12})$$

where the last line follows because

$$\begin{aligned} \theta_m(k, r, u, s, v) &:= \frac{1}{m} \tau^{k^2(us-rv)} \text{tr}(U_{r-u,s-v}^k) \\ &= \begin{cases} 1 & \text{if } k \bmod m = 0 \text{ or both } r = u \text{ and } s = v \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (\text{C13})$$

The reason that this proof works only for odd prime dimensions is that, if  $m$  is not prime, then there will exist a  $k$  and  $a, b, c, e$  such that the lhs of Eq. (C13) is 1 for conditions other than those of Eq. (C13). Furthermore, when  $m = 2$ , then there will be diagonal elements such that Eq. (C13) is  $-1$  for  $k = 2$ , and only for  $k = 4$  do we get actual recurrence (implying in turn that for  $m = 2$  the map is neither the dephasing map nor the identity map).

However, in the following lemma, we show that for any odd dimension we can construct a unitary operator basis that does exhibit recurrence.

*Lemma 10 (Recurrence for odd dimension).*—Let  $\dim \mathcal{H}_S = m^2$ ,  $\dim \mathcal{H}_R = m$ , where  $m$  is odd. There exists a unitary  $V$  acting on  $\mathcal{H}_S \otimes \mathcal{H}_R$  such that

$$\text{tr}_B[V^k(\rho \otimes \mathbb{I}_m)(V^\dagger)^k] = \pi_m^k(\rho). \quad (\text{C14})$$

*Proof.*—Consider the prime factor decomposition of  $m = p_1 \dots p_l$ . We can split the Hilbert spaces as

$$\mathcal{H}_R \simeq \bigotimes_{j=1}^l \mathcal{H}_j, \quad (\text{C15})$$

where  $\dim(\mathcal{H}_j) = p_j$ . Moreover, let  $A = \{|\mathbf{r}, \mathbf{s}\rangle\}$  be an orthonormal basis of  $\mathcal{H}_S$ , where  $\mathbf{r}, \mathbf{s} \in \mathcal{S} := \times_{j=1}^l \{1, \dots, p_j\}$ , so that  $|\mathcal{S}| = m$ . Now, we define the unitary

$$V = \sum_{\mathbf{r}, \mathbf{s} \in \mathcal{S}} |\mathbf{r}, \mathbf{s}\rangle \langle \mathbf{r}, \mathbf{s}|_S \otimes \left( \bigotimes_j U_{r_j, s_j}^{(j)} \right)_R, \quad (\text{C16})$$

where  $U_{r,s}^{(j)}$  acts nontrivially only on  $\mathcal{H}_j$  and  $r_j, s_j$  denote the  $j$ th component of the respective strings. The result now follows in just the same way as in the previous proof, as

$$\mathrm{tr}_B[V^k(\rho \otimes \mathbb{I}/m)(V^\dagger)^k] = \sum_{\mathbf{r}, \mathbf{s}, \mathbf{u}, \mathbf{v}} |\mathbf{r}, \mathbf{s}\rangle \langle \mathbf{r}, \mathbf{s} | \rho | \mathbf{u}, \mathbf{v}\rangle \langle \mathbf{u}, \mathbf{v} | \prod_j^l \left( \frac{1}{p_j} \mathrm{tr}(U_{kr_j, ks_j}^{(j)} U_{-ku_j, -kv_j}^{(j)}) \right) \quad (\text{C17})$$

$$= \sum_{\mathbf{r}, \mathbf{s}, \mathbf{u}, \mathbf{v}} |\mathbf{r}, \mathbf{s}\rangle \langle \mathbf{r}, \mathbf{s} | \rho | \mathbf{u}, \mathbf{v}\rangle \langle \mathbf{u}, \mathbf{v} | \prod_j^l \theta_{p_j}(k, r_j, s_j, u_j, v_j) \quad (\text{C18})$$

$$= \pi_m^k(\rho), \quad (\text{C19})$$

since  $k = m$  is by construction the smallest integer such that  $k \bmod p_j = 0$  for all  $j$ .  $\blacksquare$

Also, it should be noted that the case of even dimension can also be considered very close to a perfect dephasing map: Within the cycle  $k \in \{1, \dots, 2m\}$ , the only two times at which the above map does not dephase perfectly is at  $k = m$  and  $k = 2m$ . At the latter, it yields the identity map, while at the former, it yields the identity map up to sign flips on a subset of its elements.

#### APPENDIX D: ENTANGLEMENT-ASSISTED PRIVATE QUANTUM CHANNEL

Here, we present the proofs for the ideal PQC presented in the main text and discuss its properties. As our construction does not fit into the usual formal framework of PQCs with classical keys, let us first specify in more detail what we mean by a private quantum channel with a quantum key. We assume that Alice and Bob hold a shared quantum system  $K = K_A K_B$  in a state vector  $|\Psi\rangle_K$ , which we refer to as the key, and that Alice wants to encode a quantum system  $S$  with Hilbert space  $\mathcal{H}_S$ . For notational simplicity, we write  $\mathcal{H}_{K_A} = \mathcal{H}_A$  and  $\mathcal{H}_{K_B} = \mathcal{H}_B$ . Then an ideal private quantum channel with key  $|\Psi\rangle_K$  is given by a pair of quantum channels  $\mathcal{X}: \mathcal{S}(\mathcal{H}_S \otimes \mathcal{H}_A) \rightarrow \mathcal{S}(\mathcal{H}'_S \otimes \mathcal{H}_A)$  and  $\mathcal{Y}: \mathcal{S}(\mathcal{H}'_S \otimes \mathcal{H}_B) \rightarrow \mathcal{S}(\mathcal{H}_S \otimes \mathcal{H}_B)$  with the following properties. First, there exists a fixed state  $\tau$ , such that for all auxiliary systems  $E$  and all states  $\rho_{SE}$  on  $S$  and  $E$ , we have

$$\mathrm{tr}_K \circ (\mathcal{X} \otimes \mathrm{id}_{K_B E})(\rho_{SE} \otimes |\Psi\rangle \langle \Psi|_K) = \tau \otimes \rho_E. \quad (\text{D1})$$

This implies that an eavesdropper cannot learn anything from the encoded message, even when previously entangled with  $S$ . Second, the transmission is reliable; that is, for all states  $\rho$  on  $S$ , we have

$$\mathrm{tr}_K \circ (\mathcal{Y} \otimes \mathrm{id}_{K_A}) \circ (\mathcal{X} \otimes \mathrm{id}_{K_B})(\rho \otimes |\Psi\rangle \langle \Psi|_K) = \rho. \quad (\text{D2})$$

In the following, we show that the construction sketched in the main text fulfills this definition and explore some of its additional properties. We begin with the following lemma.

*Lemma 11 (Properties of a private quantum channel).—* Let  $\rho \in \mathcal{S}(\mathcal{H}_S)$  with  $\dim(\mathcal{H}_S) = d$  and let  $|\phi^+\rangle \in \mathcal{H}_K = \mathcal{H}_A \otimes \mathcal{H}_B$  be an  $e$ -dimensional, maximally entangled bipartite state vector with  $e = (\lceil d^{1/2} \rceil)^2$ . Then there exist unitaries  $U \in \mathcal{B}(\mathcal{H}_S \otimes \mathcal{H}_A)$ ,  $V \in \mathcal{B}(\mathcal{H}_S \otimes \mathcal{H}_B)$  such that

$$\mathrm{tr}_{A,B}[U(\rho \otimes |\phi^+\rangle \langle \phi^+|)U^\dagger] = \mathbb{I}_d, \quad \forall \rho, \quad (\text{D3})$$

and

$$VU(\rho \otimes |\phi^+\rangle \langle \phi^+|)U^\dagger V^\dagger = \rho \otimes |\phi^+\rangle \langle \phi^+|, \quad \forall \rho. \quad (\text{D4})$$

*Proof.*—Consider first the case that  $d$  is a square number, in which case  $e = d$ . We can assume without loss of generality that

$$|\phi^+\rangle = |\phi_1^+\rangle \otimes |\phi_2^+\rangle, \quad (\text{D5})$$

where  $|\phi_i^+\rangle$  are both  $\sqrt{e}$ -dimensional maximally entangled state vectors acting on  $\mathcal{H}_{A_i} \otimes \mathcal{H}_{B_i}$ , respectively, of the form

$$|\phi_i^+\rangle = \frac{1}{e^{1/4}} \sum_{j=1}^{\sqrt{e}} |j, j\rangle_{A_i B_i}. \quad (\text{D6})$$

We can do this because Alice and Bob can always rotate between all maximally entangled states by applying local unitaries and hence prepare the above state. We now define the unitaries

$$U_I = \sum_i^d |i\rangle \langle i|_S \otimes (U_i)_{A_1}, \quad (\text{D7})$$

$$U_J = \sum_j^d |j\rangle \langle j|_S \otimes (U_j)_{A_2}, \quad (\text{D8})$$

$$U = U_J U_I, \quad (\text{D9})$$

where  $\{U_i\}_{i=1}^d, \{U_j\}_{j=1}^d$  are unitary operator bases for  $\mathcal{H}_{A_1}$  and  $\mathcal{H}_{A_2}$ , respectively, and  $I = \{|i\rangle\}_{i=1}^d$  and  $J = \{|j\rangle\}_{j=1}^d$

are any two mutually unbiased bases (MUBs) for  $\mathcal{H}_S$ ; that is, they are both orthonormal and

$$|\langle i|j\rangle|^2 = \frac{1}{d}, \quad \forall i, j. \quad (\text{D10})$$

In prime power dimension, there are known to exist sets of  $d+1$  many of such MUBs, but there exist at least two in any dimension [30].

By direct evaluation, we now have

$$\text{tr}_{A,B}[U(\rho \otimes |\phi^+\rangle\langle\phi^+|)U^\dagger] \quad (\text{D11})$$

$$= \sum_{i,i',j,j'} |j\rangle\langle j|i\rangle\langle i|\rho|i'\rangle\langle i'|j'\rangle\langle j'| \text{tr}(U_i U_{i'}) \text{tr}(U_j U_{j'}) / d \quad (\text{D12})$$

$$= \sum_j \text{tr}(\rho) \frac{1}{d} |j\rangle\langle j| = \mathbb{I}_d, \quad (\text{D13})$$

where we use both the orthonormality of the operator bases and the defining property of the MUBs.

We now turn to the unitary  $V$ . The construction is very similar to that of  $U$ . In fact, we use the fact that, for any unitary  $U$ ,

$$(U \otimes \bar{U})|\phi_i^+\rangle = |\phi_i^+\rangle, \quad (\text{D14})$$

where the bar denotes complex conjugation. We therefore define

$$V_I = \sum_i^d |i\rangle\langle i|_S \otimes (\bar{U}_i)_{B_1}, \quad (\text{D15})$$

$$V_J = \sum_j^d |j\rangle\langle j|_S \otimes (\bar{U}_j)_{B_2}, \quad (\text{D16})$$

$$V = V_I V_J, \quad (\text{D17})$$

so that the unitaries now act on Bob's half of the entanglement. Equation (D4) then follows again by straightforward evaluation.

Finally, consider the case that  $d$  is not a square number.  $e$  is by construction always the smallest square number larger than, or equal to,  $d$ , so that we can always perform the splitting in Eq. (D5) in such a way that the resulting entangled states provide sufficient local randomness to perform the two dephasing operations. ■

The above can now be used to construct an ideal PQC, as shown in the following.

*Lemma 12 (Ideal private quantum channels).*—With the notation from the previous lemma, the maps

$$\mathcal{X}(\cdot) := U(\cdot)U^\dagger, \quad (\text{D18})$$

$$\mathcal{Y}(\cdot) := V(\cdot)V^\dagger \quad (\text{D19})$$

form an ideal private quantum channel with key  $|\Psi\rangle_K = |\phi^+\rangle$ .

*Proof.*—The ideal reliability of the above construction follows immediately from Eq. (D4). The ideal security follows from the fact that every map  $\mathcal{R}$  with the property that it completely randomizes a given system,

$$\mathcal{R}(\rho) = \mathbb{I}_d, \quad \forall \rho \in \mathcal{S}(\mathcal{H}_S), \quad (\text{D20})$$

completely destroys all correlations that this system may have had with other systems [20], in the sense that, for any extension  $\rho_{SE}$  of some  $\rho$ ,

$$\|(\mathcal{R} \otimes \text{id})\rho_{SE} - \mathbb{I}_d \otimes \rho_E\|_1 = 0. \quad (\text{D21})$$

But since  $\text{tr}_K \circ \mathcal{X}$  has this property, by Eq. (D3), Eq. (D21) implies ideal security in the sense of Eq. (D1). ■

We now turn to a discussion of the properties of the above PQC. To begin with, note that it is catalytic in the sense that, in the absence of eavesdropping, the entanglement is, at the end, returned back in its original state. This follows from Eq. (D4). Especially since entanglement is commonly considered an expensive resource, this is a very appealing feature, even though it is not very robust, as we discuss in the next section.

Secondly, our PQC construction is optimal when considered as a noisy process, in the sense that it is impossible to construct an ideal PQC with less entanglement than we do, provided the global evolution is unitary. As in the case of the lower bounds for the dephasing map, discussed in Appendix A, we prove this optimality with respect to approximate PQCs, in order to show that our results are robust against slight deviations from an ideal PQC. To do so, we call, in analogy to the classical PQC, Eq. (38), a private quantum channel with key  $|\Psi\rangle_K$   $\epsilon$  reliable, if, instead of Eq. (D1), it satisfies

$$\sup_{\rho_{S,E} \in \mathcal{S}(\mathcal{H}_S \otimes \mathcal{H}_E)} \|\text{tr}_K \circ (\mathcal{X} \otimes \text{id}_{KBE})(\rho_{SE} \otimes |\Psi\rangle\langle\Psi|_K) - \tau \otimes \rho_E\|_1 \leq \epsilon. \quad (\text{D22})$$

*Lemma 13.*—Let  $(\mathcal{X}, \mathcal{Y})$  be an  $\epsilon$ -reliable private quantum channel with key  $|\Psi\rangle_K$  for a quantum system of dimension  $d$ . If  $\mathcal{X}$  is a unitary channel, then there exists an  $\epsilon_{cr}$  such that, for all  $\epsilon < \epsilon_{cr}$ ,

$$\dim(\mathcal{H}_A) \geq \max\{4, d^{1-\epsilon} e^{\epsilon/2}\}. \quad (\text{D23})$$

*Proof.*—The proof is fully analogous to the discussion of the quantum case in Appendix A. We therefore give only a sketch. We have that  $\text{tr}_{K_B}(|\Psi\rangle_K) = \mathbb{I}_{d_A}$ . Hence,  $\epsilon$  reliability

together with the fact that  $\mathcal{X} = U \cdot U^\dagger$  for some unitary operator  $U$  implies that the encoding channel on  $S$  is a quantum noisy operation  $\mathcal{E}_Q^{d_A}$  as defined in Eq. (2). This further implies that  $\tau = \mathbb{I}_d$ , since the von Neumann entropy is nondecreasing under noisy operations and the channel has to work for the input state  $\mathbb{I}_d$ . We now bound  $d_A$  by considering a specific transition. Let  $|\Psi\rangle_{SE}$  be the maximally entangled state over  $SE$ , where we choose the extension  $\mathcal{H}_E$  to be a copy of  $\mathcal{H}_S$ . For this particular transition,  $\epsilon$  reliability of the channel implies that

$$\|\mathcal{E}_Q^{d_A} \otimes \text{id}_E(|\Psi\rangle\langle\Psi|_{SE}) - \mathbb{I}_d \otimes \mathbb{I}_d\|_1 \leq \epsilon. \quad (\text{D24})$$

By Fannes's inequality, this implies

$$S[\mathcal{E}_Q^{d_A} \otimes \text{id}_E(|\Psi\rangle\langle\Psi|_{SE})] \geq \log d^2 + \epsilon \log(\epsilon/d^2). \quad (\text{D25})$$

We now consider the bipartition of the system  $SEA$  into  $SE$  and  $A$ . Using the Lieb-Araki inequality and following, from here on, exactly the same reasoning as that of Appendix A below Eq. (A8), yields the desired bound. ■

### 1. Error correction, authentication, key recycling

As noted above, a particularly convenient feature of our PQC construction is that it is catalytic. This property implies that, in the absence of eavesdropping, the quantum key can be fully recycled. However, it is of course the basic premise of cryptography that one is not guaranteed the absence of eavesdropping. It is therefore natural to ask how robust our PQC implementation is to eavesdropping, by asking the following questions. Can Alice and Bob correct errors inflicted by an eavesdropper? How well can Alice and Bob check whether eavesdropping has occurred? How much of the key can Alice and Bob reuse in case they detect eavesdropping?

In this section, we show that Alice and Bob can use additional ebits to error correct, authenticate efficiently, and recycle part of the key even when eavesdropping occurs. The results of this section are mostly a translation of the arguments and techniques of Ref. [25] applied to our protocol.

#### a. Error correction

We first turn to the question of error correction. Consider, for simplicity, the case that Alice and Bob want to transmit a pure 2-qubit state vector  $|\phi\rangle$  along our PQC construction (i.e., the setting depicted in Fig. 5). Following the results in the previous section,  $|\phi\rangle$  can be sent using two ebits in the Bell state vector,

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle), \quad (\text{D26})$$

as a key. We consider the effect of any Pauli error  $P_i \in \{1, X, Y, Z\}^{\otimes 2}$  that may have occurred during transmission

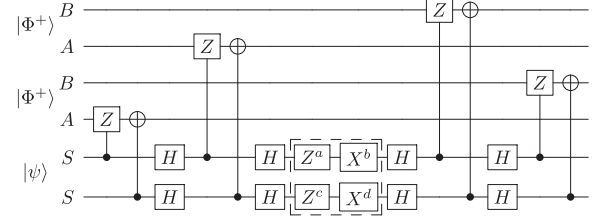


FIG. 6. The full entanglement-assisted PQC for a 2-qubit state with Pauli matrices chosen as unitary operator basis and dephasing in the computational and Pauli X eigenbases.

of the data. The reason for this is that the most general effect of eavesdropping on the encoded state  $\mathbb{I}_d = \text{tr}_K \circ \mathcal{X}(|\phi\rangle\langle\phi|)$  that is sent between Alice and Bob can be described by a quantum channel  $\mathcal{E}$  with decomposition

$$\mathcal{E}(\rho') = \sum_{i,j=0}^{15} e_{i,j} P_i \rho' P_j^\dagger. \quad (\text{D27})$$

Hence, if there exists a measurement using local operations with classical communication (LOCC) that lets Alice and Bob perfectly distinguish between any two Pauli errors without destroying the state, then they can decorrelate the message from an eavesdropper and also error correct the message [25].

We now turn to show that there exist choices for the unitary operator basis and MUBs in the PQC of Lemma 12 such that Alice and Bob can discriminate any two Pauli error without destroying the transmitted state. This possibility arises because Alice and Bob can choose the encoding in such a way that there exists a one-to-one correspondence between Pauli errors and the final state of the entanglement they used for transmission. For this correspondence to arise it suffices to (a) use the unitary operator basis defined in Eq. (C1) as bases  $\{U_i\}$  and  $\{U_j\}$  in the construction of the unitaries  $U$  and  $V$  and (b) choose  $I = \{|0\rangle, |1\rangle\}$  and  $J = \{|+\rangle = H|0\rangle, |-\rangle = H|1\rangle\}$ , where  $H$  is the Hadamard gate. For these choices, the total transmission process is given by Fig. 6, as a circuit diagram. Here, possible errors are given by the dashed box, with Alice's encoding to the left and Bob's decoding to the right of the dashed box and where we ignore global phases (for example, identifying  $Y \equiv XZ$ ) since they do not alter the outcome.

Using the relations

$$\begin{array}{c} \oplus \\ | \\ \oplus \end{array} \begin{array}{c} Z^a \\ X^b \end{array} \begin{array}{c} \oplus \\ | \\ \oplus \end{array} = \begin{array}{c} Z^a \\ X^{b+d} \end{array} \begin{array}{c} \oplus \\ | \\ \oplus \end{array} \\ \begin{array}{c} \oplus \\ | \\ \oplus \end{array} \begin{array}{c} Z^c \\ X^d \end{array} \begin{array}{c} \oplus \\ | \\ \oplus \end{array} = \begin{array}{c} Z^{a+c} \\ X^d \end{array} \begin{array}{c} \oplus \\ | \\ \oplus \end{array}$$

and



$$= (-1)^{bd}$$

$$\begin{aligned}
|\Phi^+\rangle_A |\Phi^+\rangle_S &\rightarrow |\Phi^+\rangle_A |\Phi^+\rangle_S, \\
|\Phi^+\rangle_A |\Phi^-\rangle_S &\rightarrow |\Phi^-\rangle_A |\Phi^-\rangle_S, \\
|\Phi^+\rangle_A |\Psi^+\rangle_S &\rightarrow |\Phi^+\rangle_A |\Psi^+\rangle_S, \\
|\Phi^+\rangle_A |\Psi^-\rangle_S &\rightarrow |\Phi^-\rangle_A |\Psi^-\rangle_S.
\end{aligned} \tag{D31}$$

together with the properties of entangled states, we find that a Pauli error  $P(a, b, c, d)$  described by the tuple  $(a, b, c, d) \in \{0, 1\}^{\times 4}$  yields the final state vector,

$$(X^c Z^a)_{A_1} \otimes (X^{a+d} Z^b)_{A_2} |\Phi^+\rangle_{A_1 B_1} |\Phi^+\rangle_{A_2 B_2} P(a, b, c, d) |\psi\rangle, \tag{D28}$$

ignoring global phases and omitting identity operators. This implies that we can identify the tuple  $(a, b, c, d)$  exactly just by distinguishing the Bell states, since no two different Pauli errors produce the same pair of Bell states, establishing the required correspondence. The same holds true also for mixed state messages, by linearity of quantum mechanics, and it also straightforwardly generalizes to the case of larger messages, since we can think of such messages as being sent in chunks of size 2 using the above procedure.

Going back to the case  $n = 2$ , the above establishes a one-to-one correspondence between the 16 possible Pauli errors on the ciphertext and the 16 possible combinations of Bell state vectors:

$$|\Phi^\pm\rangle = \frac{1}{\sqrt{2}}(|0, 0\rangle \pm |1, 1\rangle), \tag{D29}$$

$$|\Psi^\pm\rangle = \frac{1}{\sqrt{2}}(|0, 1\rangle \pm |1, 0\rangle). \tag{D30}$$

If Alice and Bob could discriminate between these 16 combinations using LOCC measurements, then by the above this would mean that they can both decorrelate the decoded state from an eavesdropper as well as perform error correction. However, this is not possible without the help of additional entanglement, since it is already impossible to discriminate between the four Bell states of a single ebit using LOCC measurements without further resources [31]. However, the situation is different if Alice and Bob have access to additional ebits. In particular, let  $|\chi\rangle \in \{|\Phi^\pm\rangle, |\Psi^\pm\rangle\}$  be an unknown Bell state vector. Then, if Alice and Bob share an auxiliary ebit prepared in the state vector  $|\Phi^+\rangle$ , they can each apply a CNOT gate, controlling on the auxiliary system  $A$  and targeting  $S$ , which has the effect

If Alice and Bob now each measure their share of  $A$  in the Pauli  $X$  basis and their share of  $S$  in the Pauli  $Z$  basis and broadcast their measurement results, they can perfectly identify  $|\chi\rangle$ . Using this procedure for both ebits, they can extract full information about the error on the system and correct accordingly.

In summary, we have shown that Alice and Bob can perfectly discriminate between any two Pauli errors inflicted on the ciphertext during transmission, with the help of additional  $n$  ebits. In this way, however, our PQC construction loses the advantage in resources over that of Ref. [25], where error correction is also possible using  $2n$  ebits in total.

### b. Authentication and key recycling

The above error-correcting procedure has two disadvantages: Firstly, it requires a doubling of the total entanglement and, secondly, all the entanglement gets destroyed in the process. A more resource-effective strategy of Alice and Bob is to attempt to check for the occurrence of eavesdropping, destroying as little entanglement as possible, and consequently repeat the sending of the message while reusing as much of the entanglement as possible. We now discuss such a strategy in the asymptotic case, that is, when Alice and Bob send an  $n$ -qubit quantum message  $\rho_S$  using  $n$  ebits, in the limit  $n \rightarrow \infty$ .

Let  $\mathbf{v}$  be a  $2n$ -bit string encoding the final state of the  $n$  ebits, with

$$|\Phi^+\rangle \rightarrow 00, \quad |\Psi^+\rangle \rightarrow 01, \quad |\Phi^-\rangle \rightarrow 10, \quad |\Psi^-\rangle \rightarrow 11,$$

and the first two bits corresponding to the first ebit, etc. In order to check for the occurrence of eavesdropping, Alice and Bob can employ a LOCC protocol constructed in Ref. [32] that yields the parity of any substring in  $\mathbf{v}$ , by destroying only a single ebit. Applying this protocol to  $r$  random substrings of  $\mathbf{v}$ , one has

$$\text{Prob}(\mathbf{v} \neq 00\dots 00 | \text{even parity in all } r \text{ rounds}) = \frac{1}{2^{-r}}.$$

Since  $\mathbf{v} = 00\dots 00$  corresponds to the case in which no Pauli error occurred, this implies that in case Alice and Bob measure no odd parity, they know that the message has been successfully transferred and that they can reuse their ebits for future communication, with exponentially small probability of mistake and at the cost of vanishingly few ebits. Now, in case they detect odd parity for any of their  $r$  rounds, Alice and Bob consider the transfer unsuccessful

and attempt to recycle as many of their ebits as possible. This amounts to estimating  $\mathbf{v}$  while destroying as few ebits as possible in the course of doing so. We can directly apply a key recycling procedure presented in Ref. [25] to our construction to achieve an asymptotic key recycling rate of  $[1 - H(\delta)]$ , where  $H$  is the binary Shannon entropy and  $\delta > 0$  is a security parameter. We refer the reader to Ref. [25] for details.

These results should be compared with key recycling rates for the case of classical keys. There, the achievable recycling rates depend strongly on whether the message to be sent is classical (see, e.g., Refs. [33–35]) or quantum (see, e.g., Refs. [21,22,36]), and also on the possible attack scenarios that are being considered (see Refs. [21,22] for a discussion). Overall, however, the recycling rates can be considerably higher than those obtained here, albeit with significantly more complicated authentication schemes. Improving the recycling rates in the case of quantum keys thus remains an interesting open problem.

### APPENDIX E: QUANTUM EXPANDERS

In this appendix, we discuss efficient approximate pinching to the main diagonal of an  $d$ -dimensional quantum system, of suitable dimension  $d$ , and provide the proof of Theorem 3. The proof of this statement is rooted in insights into random walks on expander graphs, is connected to properties of Wigner functions of discrete Weyl systems, and makes use of basic properties of quantum channels. It starts from and builds upon the construction presented in Ref. [27], which in turn derives from the classical description in Ref. [29]. The latter work discusses a random walk on an expander graph featuring the vertex set  $\mathbb{Z}_e^2$ , so an  $e \times e$  integer lattice. Reference [29] continues to show that the random walk it constructs converges exponentially quickly to the uniform distribution  $\mathbf{1}_{\mathbb{Z}_e^2}$  on this vertex set. Specifically, it is shown that there exists a doubly stochastic matrix such that for any probability distribution  $P$  on  $\mathbb{Z}_e^2$ , one has

$$\|S^k(P) - \mathbf{1}_{\mathbb{Z}_e^2}\|_2 \leq \frac{5\sqrt{2}}{8} \|S^{k-1}(P) - \mathbf{1}_{\mathbb{Z}_e^2}\|_2, \quad (\text{E1})$$

for  $k \geq 1$  being an integer. Here, the action of the doubly stochastic map acting upon a distribution on  $\mathbb{Z}_e^2$  is written as  $S(P)$ . On  $v = (v_p, v_q)^T \in \mathbb{Z}_e^2$ , this doubly stochastic matrix originates from random affine transformations, drawn uniformly from the following eight transformations,

$$v \mapsto T_1 v, \quad v \mapsto T_2 v, \quad v \mapsto T_1 v + e_1, \quad v \mapsto T_2 v + e_2, \quad (\text{E2})$$

and the four inverse transformations, with

$$T_1 := \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \quad T_2 := \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, \quad (\text{E3})$$

and

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (\text{E4})$$

The graph underlying this construction, with the  $e \times e$  lattice as vertex set, is an expander graph. Such an expander graph is usually referred to as an  $(e^2, 8, \lambda)$  expander graph with  $\lambda \leq 5\sqrt{2}/8$ , in that it has  $e^2$  vertices, each of which having 8 neighbors in the graph. The matrix  $S$  is sparse in that each row has 8 entries only. Clearly, the above implies that

$$\|S^k(P) - \mathbf{1}_{\mathbb{Z}_e^2}\|_2 \leq \sqrt{2} \left( \frac{5\sqrt{2}}{8} \right)^k. \quad (\text{E5})$$

The prefactor of  $\sqrt{2}$  originates from the fact that for any probability distribution  $P$  on  $\mathbb{Z}_e^2$ , one has

$$\begin{aligned} \|P - \mathbf{1}_{\mathbb{Z}_e^2}\|_2 &\leq [(1 - 1/e^2) - (e^2 - 1)/e^2]^{1/2} \\ &= \sqrt{2}(e^2 - 1)/e^2 \\ &\leq \sqrt{2}. \end{aligned} \quad (\text{E6})$$

We relate this dimension  $e$ , which is left open at this point, to the physical dimension  $d$  of the quantum system subsequently.

The construction in a significantly altered setting will require some preparation. For this, we turn to discussing the phase space  $d \times d$  for the  $d$ -dimensional quantum system with odd  $d$ . In the convention of Refs. [27,37], for phase space coordinates  $(p, q) \in \mathbb{Z}_d^2$ , the discrete Wigner function  $W_M$  of an operator  $M$  acting in Hilbert space can be written as

$$W_M(p, q) = \frac{1}{d} \text{tr}[w(p, q) \Pi w(p, q)^\dagger M], \quad (\text{E7})$$

where  $(p, q) \mapsto w(p, q)$  is the family of Weyl operators and  $\Pi$  is the parity operator. The Weyl operators are composed of shift and clock operators, so the  $X$  and  $Z$  generalized Pauli matrices defined in Eqs. (19) and (22), respectively. Any affine transformation  $A$ , the linear part of which having a unit determinant on phase space coordinates  $a \in \mathbb{Z}_d^2$ , is unitarily reflected in Hilbert space as

$$W_{U_A \rho U_A^\dagger}(a) = W_\rho(A^{-1}(a)). \quad (\text{E8})$$

Wigner functions are normalized as

$$\sum_{(p,q) \in \mathbb{Z}_d^2} W_\rho(p,q) = 1 \quad (\text{E9})$$

for quantum states  $\rho$ . We treat Wigner functions for an operator  $M$  as matrices  $W_M \in \mathbb{C}^{d \times d}$ , with real-valued matrices for Hermitian  $M$ . A first well-known insight is stated here as a separate lemma for completeness.

*Lemma 14 (Quantum states and Wigner functions).*—For two quantum states  $\rho$  and  $\sigma$  on a Hilbert space  $\mathcal{H}_S$  of dimension  $d$  associated with Wigner functions  $W_\rho, W_\sigma: \mathbb{Z}_d^2 \rightarrow \mathbb{R}$ , one has

$$\|\rho - \sigma\|_2^2 = \|W_\rho - W_\sigma\|_2^2 = \sum_{(p,q) \in \mathbb{Z}_d^2} [W_\rho(p,q) - W_\sigma(p,q)]^2. \quad (\text{E10})$$

*Proof.*—This statement follows directly from the property that the Hilbert-Schmidt scalar product is inherited as

$$\text{tr}(\rho\sigma) = \sum_{(p,q) \in \mathbb{Z}_d^2} W_\rho(p,q)W_\sigma(p,q), \quad (\text{E11})$$

as follows from the analogous property of the characteristic function, and the definition of the 2-norm. ■

The main insight of Ref. [27] is to acknowledge that random walks on integer lattices that are expander graphs can be connected to random unitary channels acting in Hilbert space that inherit the mixing properties from the classical random walk, by resorting to a phase space picture. The construction of Ref. [27] builds upon the random walk on the Margulis expander graph [29], the vertex set of which is  $\mathbb{Z}_e^2$  for some  $e$  (here taken to be different from  $d$ , as it will take a different role subsequently). This random walk can be unitarily realized in quantum systems: In fact, the random walk follows directly from a convergence of a Wigner function, a function that shares all properties of a probability distribution, except being positive. Following the construction of the random walk on the expander graph, the quantum Margulis expander can be seen as a random unitary map,

$$\rho \mapsto \mathcal{D}(\rho) = \frac{1}{8} \sum_{j=1}^8 U_j \rho U_j^\dagger, \quad (\text{E12})$$

of Kraus rank 8 with suitable unitary  $\{U_i\}$  with the property that

$$\|\mathcal{D}(\rho) - \mathbb{I}_e\|_2 \leq \frac{5\sqrt{2}}{8} \|\rho - \mathbb{I}_e\|_2. \quad (\text{E13})$$

A second insight on discrete Wigner functions that we will make use of is the following.

*Lemma 15 (Wigner functions of pinched quantum states).*—For any quantum state  $\rho$  on  $\mathcal{H}_S$  of dimension  $d$ , the Wigner function of  $\pi(\rho)$  satisfies

$$W_{\pi(\rho)}(p,q) = W_{\pi(\rho)}(p',q), \quad (\text{E14})$$

for all  $q, p, p' = 1, \dots, d$ .

*Proof.*—This statement follows directly from the definition of Wigner functions. ■

This means that Wigner functions of pinched states are constant along the first coordinate. Prepared in this fashion, we can finally turn to the new construction. This construction of a random unitary channel will deviate from this construction in a significant way: We identify for each  $q \in \mathbb{Z}_d$  for  $d = e^2$  the entire line  $\{(p,q) \in \mathbb{Z}_d^2\}$  of the  $(d \times d)$ -dimensional phase space as a vectorized  $e \times e$  lattice, on which the above affine maps act. The property of the unit determinant of the linear part in the affine mapping is preserved. In fact, it will act in precisely the same way on each line simultaneously, by applying one of the 8 affine transformations defined in Eqs. (E2)–(E4). This gives rise to 8 affine maps on  $\mathbb{Z}_d^2$ . Acting on Wigner functions, this process can again be realized as a random unitary channel,

$$\rho \mapsto \mathcal{T}(\rho) = \frac{1}{8} \sum_{j=1}^8 V_j \rho V_j^\dagger, \quad (\text{E15})$$

with unitaries  $\{V_i\}$ . Clearly, the entire Wigner function  $W_\rho$  of a state is normalized according to Eq. (E9). We refer to

$$x_q := \sum_{p \in \mathbb{Z}_d} W_\rho(p,q) \quad (\text{E16})$$

as the weight of each column. We now discuss the convergence properties of the above random unitary channel. For an integer  $k \geq 1$ , we have

$$\begin{aligned} \|\mathcal{T}^k(\rho) - \pi(\rho)\|_2^2 &= \|W_{\mathcal{T}^k(\rho)} - W_{\pi(\rho)}\|_2^2 \\ &= \sum_{q \in \mathbb{Z}_d} x_q^2 \sum_{p \in \mathbb{Z}_d} \left( \frac{W_{\mathcal{T}^k(\rho)}(p,q)}{x_q} - \frac{1}{d} \right)^2, \end{aligned} \quad (\text{E17})$$

treating each columns separately. Using  $x_q \leq d$  for all  $q$  and using a worst-case bound for all  $q$  gives

$$\|\mathcal{T}^k(\rho) - \pi(\rho)\|_2^2 \leq d^3 \sum_{p \in \mathbb{Z}_d} \left( \frac{W_{\mathcal{T}^k(\rho)}(p,q)}{x_q} - \frac{1}{d} \right)^2, \quad (\text{E18})$$

and following Eq. (E5), one obtains

$$\|\mathcal{T}^k(\rho) - \pi(\rho)\|_2^2 \leq 2d^3 \left( \frac{5\sqrt{2}}{8} \right)^{2k}. \quad (\text{E19})$$

In this way, we arrive at the anticipated result, by embedding the random unitary system into an explicit quantum model, in the nomenclature of the main text.



- [1] G. Gour, M. P. Mueller, V. Narasimhachar, R. W. Spekkens, and N. Y. Halpern, *The Resource Theory of Informational Nonequilibrium in Thermodynamics*, *Phys. Rep.* **583**, 1 (2015).
- [2] C. Gogolin and J. Eisert, *Equilibration, Thermalisation, and the Emergence of Statistical Mechanics in Closed Quantum Systems*, *Rep. Prog. Phys.* **79**, 056001 (2016).
- [3] N. Linden, S. Popescu, A. J. Short, and A. Winter, *Quantum Mechanical Evolution towards Thermal Equilibrium*, *Phys. Rev. E* **79**, 061103 (2009).
- [4] W. H. Zurek, *Decoherence, Einselection, and the Quantum Origins of the Classical*, *Rev. Mod. Phys.* **75**, 715 (2003).
- [5] E. Joos, H. D. Zeh, C. Kiefer, D. J. W. Giulini, J. Kupsch, and I.-O. Stamatescu, *Decoherence and the Appearance of a Classical World in Quantum Theory* (Springer, Berlin, 2003).
- [6] A. S. Holevo, *Statistical Structure of Quantum Theory* (Springer, Berlin, 2001).
- [7] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2000).
- [8] A. Ambainis, M. Mosca, A. Tapp, and R. De Wolf, *Private Quantum Channels*, in *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, (IEEE Computer Society, 2000), pp. 547–553.
- [9] P. O. Boykin and V. Roychowdhury, *Optimal Encryption of Quantum Bits*, *Phys. Rev. A* **67**, 042317 (2003).
- [10] M. P. Mueller, *Correlating Thermal Machines and the Second Law at the Nanoscale*, [arXiv:1707.03451](https://arxiv.org/abs/1707.03451).
- [11] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications* (Springer, New York, 2011).
- [12] A. Horn, *Doubly Stochastic Matrices and the Diagonal of a Rotation Matrix*, *Am. J. Math.* **76**, 620 (1954).
- [13] J. Scharlau and M. P. Mueller, *Quantum Horn’s Lemma, Finite Heat Baths, and the Third Law of Thermodynamics*, *Quantum* **2**, 54 (2018).
- [14] J. Schwinger, *Unitary Operator Bases*, *Proc. Natl. Acad. Sci. U.S.A.* **46**, 570 (1960).
- [15] R. F. Werner, *All Teleportation and Dense Coding Schemes*, *J. Phys. A* **34**, 7081 (2001).
- [16] D. Jonathan and M. B. Plenio, *Entanglement-Assisted Local Manipulation of Pure Quantum States*, *Phys. Rev. Lett.* **83**, 1455 (1999).
- [17] N. H. Y. Ng, L. Mancinska, C. Cirstoiu, J. Eisert, and S. Wehner, *Limits to Catalysis in Quantum Thermodynamics*, *New J. Phys.* **17**, 085004 (2015).
- [18] J. Eisert and M. Wilkens, *Catalysis of Entanglement Manipulation for Mixed States*, *Phys. Rev. Lett.* **85**, 437 (2000).
- [19] A. Ambainis, J. Bouda, and A. Winter, *Nonmalleable Encryption of Quantum Information*, *J. Math. Phys. (N.Y.)* **50**, 042106 (2009).
- [20] P. Hayden, D. Leung, P. W. Shor, and A. Winter, *Randomizing Quantum States: Constructions and Applications*, *Commun. Math. Phys.* **250**, 371 (2004).
- [21] C. Portmann, *Quantum Authentication with Key Recycling*, in *Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques* (Springer, New York, 2017), pp. 339–368.
- [22] P. Hayden, D. W. Leung, and D. Mayers, *The Universal Composable Security of Quantum Message Authentication with Key Recycling*, [arXiv:1610.09434](https://arxiv.org/abs/1610.09434).
- [23] If instead of the diamond norm a PQC’s security is defined with respect to the weaker trace norm, then any unitary 1-design provides an ideal channel and there exist both randomized [20] and deterministic [24] constructions of PQCs that require only  $n + O[\log(n)]$  many bits of shared key.
- [24] A. Ambainis and A. Smith, *Small Pseudo-Random Families of Matrices: Derandomizing Approximate Quantum Encryption*, in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques. RANDOM 2004, APPROX 2004*, Lecture Notes in Computer Science (Springer, Berlin, Heidelberg, 2004).
- [25] D. W. Leung, *Quantum Vernam Cipher*, *Quantum Inf. Comput.* **2**, 14 (2002).
- [26] M. B. Hastings, *Random Unitaries Give Quantum Expanders*, *Phys. Rev. A* **76**, 032315 (2007).
- [27] D. Gross and J. Eisert, *Quantum Margulis Expanders*, *Quantum Inf. Comput.* **8**, 722 (2008).
- [28] A. W. Harrow, *Quantum Expanders from Any Classical Cayley Graph Expander*, *Quantum Inf. Comput.* **8**, 715 (2008).
- [29] G. A. Margulis, *Explicit Constructions of Concentrators*, *Prob. Peredachi Inf.* **9**, 71 (1973).
- [30] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement*, 2nd ed. (Cambridge University Press, Cambridge, England, 2017).
- [31] J. Walgate, A. J. Short, L. Hardy, and V. Vedral, *Local Distinguishability of Multipartite Orthogonal Quantum States*, *Phys. Rev. Lett.* **85**, 4972 (2000).
- [32] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, *Mixed-State Entanglement and Quantum Error Correction*, *Phys. Rev. A* **54**, 3824 (1996).
- [33] C. H. Bennett, G. Brassard, and S. Breidbart, *Quantum Cryptography II: How to Re-Use a One-Time Pad Safely Even if P = NP*, *Nat. Comput.* **13**, 453 (2014).
- [34] I. Damgård, T. B. Pedersen, and L. Salvail, *A quantum cipher with near optimal key-recycling*, in *CRYPTO 2005: 25th Annual International Cryptology Conference, Santa Barbara*, Lecture Notes in Computer Science, Vol. 3621 (Springer, New York, 2005), pp. 494–510.
- [35] S. Fehr and L. Salvail, *Quantum Authentication and Encryption with Key Recycling*, in *Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques* (Springer, New York, 2017), pp. 311–338.
- [36] J. Oppenheim and M. Horodecki, *How to Reuse a One-Time Pad and Other Notes on Authentication, Encryption, and Protection of Quantum Information*, *Phys. Rev. A* **72**, 042309 (2005).
- [37] D. Gross, *Hudson’s Theorem for Finite-Dimensional Quantum Systems*, *J. Math. Phys. (N.Y.)* **47**, 122107 (2006).



## CORRELATED CATALYSIS

## 5.1 MOTIVATION AND CHARACTERIZATION

## 5.1.1 Reusing catalysts on sequences of systems

In the previous chapter, one particularly interesting property of the dephasing construction that we introduced was the fact that it leaves the reduced state of the source of randomness unchanged. That is, for any pair of states  $\rho, \rho'$  that act on a  $d$ -dimensional Hilbert space and that satisfy  $\rho \succeq \rho'$ , we constructed a unitary  $U$  such that

$$\mathrm{Tr}_R \left[ U(\rho \otimes \mathbf{1}_{\lceil \sqrt{d} \rceil}) U^\dagger \right] = \rho'$$

while

$$\mathrm{Tr}_S \left[ U(\rho \otimes \mathbf{1}_{\lceil \sqrt{d} \rceil}) U^\dagger \right] = \mathbf{1}_{\lceil \sqrt{d} \rceil}.$$

Operationally, this property has the interesting implication that one can re-use the system  $R$  to act as a source of randomness for third systems. That is, let  $(S_i)_i$  be a sequence of  $d$ -dimensional systems, and let  $(\rho_i)_i, (\rho'_i)_i$  be sequences of their states on  $S_i$  such that  $\rho_i \succeq \rho'_i$  for every  $i$ . Then, by the above property of the construction of  $U$ , we can use a *single* maximally mixed state of dimension  $\lceil \sqrt{d} \rceil$  to act as the source of randomness that realizes all transitions  $\rho_i \rightarrow \rho'_i$ . More precisely, the above guarantees that there exists a sequence of unitary channels  $(\mathcal{U}_i)_i$  where each unitary channel has support only on  $S_i$  and  $R$ , such that the state

$$\rho' = \cdots \circ \mathcal{U}_2 \circ \mathcal{U}_1 (\mathbf{1}_{\lceil \sqrt{d} \rceil} \otimes (\bigotimes_i \rho_i))$$

satisfies

$$\mathrm{Tr}_{S_i^c}[\rho'] = \rho'_i, \quad \mathrm{Tr}_{R^c}[\mathbf{1}_{\lceil \sqrt{d} \rceil}] = \mathrm{Tr}_{R^c}[\mathbf{1}_{\lceil \sqrt{d} \rceil}].$$

Note that the above reasoning is independent of the length of the sequence. Now, in principle it is a non-trivial constraint on a noisy operation to require that the reduced state of the source of randomness remains unchanged by the joint unitary acting on  $S$  and  $R$ , because the definition of a noisy operation makes no requirement on the final state of  $R$ . So it is interesting to observe that adding this constraint to a noisy operation – that the operation should leave the state of  $R$  locally unchanged – does not in fact put any further constraint on the possible transitions: Majorization, which characterizes possible state transitions under noisy operations, also characterizes the state transitions that are possible under the restricted subset of noisy operations that have to satisfy this additional constraint.

The key idea leading to the topic of this chapter is that the operational advantage of having a joint unitary that leaves the marginal state of the ancilla  $R$  unchanged, namely to reuse the ancilla with additional system, does not hinge on the fact that the initial state of this ancilla is maximally mixed. That is, if one could realise interesting state transitions on  $S$  by means of a unitary and an initial state  $\sigma$  on  $R$  such that the final state on  $R$  would again be  $\sigma$ , then one could reuse a single copy of  $R$  to realise those interesting state transitions on an arbitrary number of systems in just the same way as described above. The study of state transitions

that can be realized in this way is the study of the power of *correlated catalysis*. Recall from Sec. 3.3 the notion of uncorrelated catalysis. There, one requires not only that the state  $\sigma$  of an ancilla is returned locally unchanged, but that it moreover be uncorrelated from the system  $S$  at the end of the transition. Here, in contrast, we only require local invariance of the ancilla but not that the systems be uncorrelated at the end. This is because the ability of the catalyst to locally enable transitions on some system  $S_i$  is *independent* of the question whether the catalyst is correlated with some other system  $S_j$ . Of course, globally, whether correlations between systems and the catalyst are established will affect the correlations that are established *between* systems in the course of interacting with the catalyst  $C$ . That is, if  $(\mathcal{U}_i)_i$  and  $(\rho_i)_i$  denote sequences of unitary channels and states respectively such that  $\mathcal{U}_i$  acts on  $S_i$  and a catalyst  $C$  that is initially in state  $\sigma$  and left invariant under  $\mathcal{U}_i$ , then if

$$\mathcal{U}_i(\rho_i \otimes \sigma) = \rho'_i \otimes \sigma, \quad \forall i,$$

this implies that

$$\mathrm{Tr}_C [\cdots \circ \mathcal{U}_2 \circ \mathcal{U}_1 (\sigma \otimes (\otimes_i \rho_i))] = \bigotimes_i \rho'_i,$$

i.e. the marginal state on the systems  $(S_i)_i$  is also uncorrelated. This is in contrast with correlated catalysis, where I can only infer that the transitions has occurred at each  $S_i$  locally (as in Eq. 5.1.1), but not that the final states on two different systems are uncorrelated. But in settings in which such correlations are not problematic, the usual requirement of uncorrelated catalysis seems operationally unnecessarily strong.<sup>1</sup> This could be the case, for example, if the systems  $S_i$  are guaranteed never to interact with another. Indeed, as we will see below, there might even be situations in which these correlations are advantageous for a task.

### 5.1.2 Characterization via von Neumann entropy

What are the state transitions that are possible under correlated catalysis? This is the main question of this section. The main result of the following publication [2] is that, under a minor modification of the above setting, in which the catalyst only has to be returned up to coherences in some fixed basis, the possible state transitions are characterized by the von Neumann entropy, in the sense that, roughly speaking, a transition  $\rho \rightarrow \rho'$  between two full-rank states in this slightly modified setting is possible if and only if  $S(\rho') > S(\rho)$ . This is remarkable in that it provides an operational *single-shot* interpretation of the von Neumann entropy. This is in contrast to folklore knowledge which says that the von Neumann entropy is singled out as a special monotone only in the thermodynamic limit (as in Sec. 3.2). The chapter also presents a conjecture which essentially states that the von Neumann entropy also characterizes possible state transitions in the unmodified setting of correlated catalysis. This conjecture is, to the author's knowledge, still unproven and increasingly appreciated as an important open problem in quantum information theory [102, 143], with some progress reported in [144, 145]. For instance, it is now known that for some transitions that are possible under catalytic catalysis, the dimension of the catalyst needs to become very large [3, 8].

<sup>1</sup> Indeed, from a strictly operational point, one could define an even weaker notion of catalysis, in which the catalyst can change its local state, as long as its ability to assist in a given state transition is not compromised. While tricky to define in full generality, for sequences of fixed transitions  $\rho \rightarrow \rho'$  this can be done and, interestingly be shown to coincide in power to the notion of correlated catalysis considered here [142].

## Von Neumann Entropy from Unitarity

Paul Boes,<sup>1</sup> Jens Eisert,<sup>1</sup> Rodrigo Gallego,<sup>1</sup> Markus P. Müller,<sup>2,3</sup> and Henrik Wilming<sup>4</sup>

<sup>1</sup>*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*

<sup>2</sup>*Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Boltzmannngasse 3, A-1090 Vienna, Austria*

<sup>3</sup>*Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*

<sup>4</sup>*Institute for Theoretical Physics, ETH Zurich, 8093 Zurich, Switzerland*



(Received 16 August 2018; revised manuscript received 26 March 2019; published 28 May 2019)

The von Neumann entropy is a key quantity in quantum information theory and, roughly speaking, quantifies the amount of quantum information contained in a state when many identical and independent (i.i.d.) copies of the state are available, in a regime that is often referred to as being asymptotic. In this Letter, we provide a new operational characterization of the von Neumann entropy which neither requires an i.i.d. limit nor any explicit randomness. We do so by showing that the von Neumann entropy fully characterizes single-shot state transitions in unitary quantum mechanics, as long as one has access to a catalyst—an ancillary system that can be reused after the transition—and an environment which has the effect of dephasing in a preferred basis. Building upon these insights, we formulate and provide evidence for the *catalytic entropy conjecture*, which states that the above result holds true even in the absence of decoherence. If true, this would prove an intimate connection between single-shot state transitions in unitary quantum mechanics and the von Neumann entropy. Our results add significant support to recent insights that, contrary to common wisdom, the standard von Neumann entropy also characterizes single-shot situations and opens up the possibility for operational single-shot interpretations of other standard entropic quantities. We discuss implications of these insights to readings of the third law of quantum thermodynamics and hint at potentially profound implications to holography.

DOI: [10.1103/PhysRevLett.122.210402](https://doi.org/10.1103/PhysRevLett.122.210402)

In quantum information theory, it is common to distinguish tasks as falling in one of two regimes: Either (i) one deals with situations in which many identically and independently distributed (i.i.d.) quantum systems appear. This regime is usually referred to as the *asymptotic regime*. Such tasks include, for example, Schumacher compression [1], entanglement distillation [2], and quantum hypothesis testing [3,4]. Or, in sharp contrast, (ii) one deals with situations that involve only a single quantum system, the so-called *single-shot* regime. Examples of protocols that have been analyzed in the single-shot setting include the decoupling of quantum systems [5], hypothesis testing [6], and state transitions in quantum thermodynamics [7].

Common wisdom has it that different quantities characterize these two regimes. In the first regime, the von Neumann entropy (VNE) or quantities directly related to it prevail, such as the standard quantum relative entropy or mutual information, while in the second regime quantities such as quantum Rényi divergences [8–11] and smoothed versions of the above [12,13] become important. This common wisdom is, however, recently being challenged [14–19], as it has been shown that the VNE determines possible single-shot state transitions in quantum mechanics—under unitary evolutions—provided that three assumptions hold [18]: (i) One can prepare a suitable auxiliary system that does not change its state during the

process but might become correlated with the system on which the transition is performed; (ii) one has access to an environment, or source of randomness, that is modeled as a large system in the maximally mixed state; and (iii) one has full control over the system, auxiliary system, and environment, in the sense that one can implement any unitary on the joint system. Assumption (ii) assigns an undesirably special role to maximally mixed systems, while assumption (iii) is in conflict with the common experience that environments cannot practically be accessed with a full degree of control.

In this work, we provide an operational characterization of the von Neumann entropy in terms of single-shot state transitions that, remarkably, does without assumptions (ii) and (iii). Instead, our characterization builds upon two natural classes of dynamics in quantum mechanics: controlled unitary evolution and uncontrolled decoherence to some given preferred basis. We also apply this characterization to a notion of cooling that is usually considered in the context of quantum readings of the *third law of thermodynamics* and discuss possible implications of our results for recent work on the decoupling of systems and the AdS/CFT correspondence in the context of *holography*. Finally, we formulate, and provide evidence for, a conjecture, which states that not even decoherence is necessary to single out VNE and, if true, would show that the von

Neumann entropy can be derived directly from unitary quantum mechanics alone.

*Main result.*—We will now present our main result and then discuss its implications. To state the result, let  $\mathcal{D}_J$  be the quantum channel that decoheres a system in a given orthonormal basis  $J := \{|j\rangle\}$  of its Hilbert space, according to

$$\mathcal{D}_J[\sigma] = \sum_j \langle j|\sigma|j\rangle |j\rangle\langle j|.$$

Density matrices diagonal in  $\{|j\rangle\}$  will be called *quasiclassical*. Our main result can be stated as follows.

*Theorem 1: Single-shot characterization of the von Neumann entropy.*—Let  $\rho$  and  $\rho'$  be two density matrices of the same finite dimension and with different spectra. Then the following two statements are equivalent: (i)  $S(\rho') > S(\rho)$  and  $\text{rank}(\rho') \geq \text{rank}(\rho)$ . (ii) There exists a finite-dimensional Hilbert space, for any basis  $J$  of which there exists a quasiclassical density matrix  $\sigma$  and a unitary  $U$  such that

$$\text{Tr}_2[U(\rho \otimes \sigma)U^\dagger] = \rho', \quad (1)$$

$$\mathcal{D}_J[\text{Tr}_1[U(\rho \otimes \sigma)U^\dagger]] = \sigma. \quad (2)$$

The proof is presented in Sec. I in Supplemental Material [20]. Note first that if one has  $S(\rho') > S(\rho)$  but  $\text{rank}(\rho') < \text{rank}(\rho)$ , then by Theorem 1 the transition is not possible exactly. However, it can be done to an arbitrary precision, since any state can be arbitrarily well approximated by a state with full rank. From a physical point of view, the condition on the rank is therefore not important.

To interpret this result, one can imagine a situation in which only a small region of space, say, the laboratory, can be controlled unitarily with a high degree of precision while any system outside this region is decohered very quickly in some given basis. This is a common situation in current experimental devices. Given these constraints, the goal is to transform a quantum system from  $\rho$  to  $\rho'$  by acting unitarily on this system together with an ancillary system in a quasiclassical state that one can “borrow” from the environment so long as, upon being returned to the environment, it decoheres back to its initial state and can hence be used to aid further transitions. Then, Theorem 1 says that the VNE fully characterizes possible transitions in this natural setup.

In general, the auxiliary system is clearly necessary to implement the transition  $\rho \rightarrow \rho'$ , since, otherwise, we would act unitarily on  $\rho$  and, therefore, could not change its spectrum. The same restriction would arise if we demanded that the auxiliary system is returned *uncorrelated* from the system. Finally, it can be reused to enable further transitions  $\rho \rightarrow \rho'$  on independent copies of  $\rho$ . This is true even if correlations are established between the

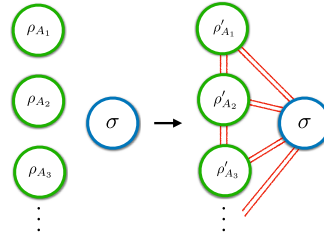


FIG. 1. Reusability of the auxiliary system for further transitions. Consider  $N$  subsystems in an uncorrelated state  $\rho_{A_1, \dots, A_N} = \rho^{\otimes N}$ . Because of Theorem 1, for any transition  $\rho_{A_1} \rightarrow \rho'_{A_1}$  respecting the entropy and rank condition, it is possible to find an auxiliary system in state  $\sigma$  that enables this transition. When brought back in contact with the environment, it dephases and returns to its initial state while establishing correlations with  $A_1$ . In spite of these correlations, it is reusable to implement the same transition on  $A_2$ . This is true, since in the second step one applies a local operation on  $A_2$  and the auxiliary system, whose outcome is independent of the correlations with  $A_1$ . Repeating this process on the  $N$  subsystems results in having performed locally transitions  $\{\rho_{A_i} \rightarrow \rho'_{A_i}\}_{i=1, \dots, N}$  while using a single auxiliary system. At the end of the process, all the subsystems are possibly correlated. However, these correlations do not play any role if one intends to use each subsystem  $A_i$  independently for further thermodynamic or information protocols, as is generally the case in a single-shot setting.

auxiliary system and the system of interest in each transition (see Fig. 1).

Thus, the auxiliary system acts like a *catalyst*, in the sense that it enables transitions that would otherwise be impossible without being degraded itself. The notion of catalysis we employ here, however, is different from the one commonly used in resource theories, where the catalyst is usually required to be returned uncorrelated to the system of interest (but may become correlated to other systems, e.g., heat baths in quantum thermodynamics). Finally, we emphasize that, as is usual for catalysts, the auxiliary system and its state  $\sigma$  depend on the transition  $\rho \rightarrow \rho'$  and on the dephasing basis, which we think of as being determined by the environment (and, hence, can be expected to coincide with the energy eigenbasis).

*Applications to notions of cooling and the third law.*—We now discuss an application of Theorem 1 to one of the key problems in quantum thermodynamics. Namely, we analyze how it can be used as a protocol for *cooling to very low temperatures* beyond the i.i.d. setting. This is a situation usually captured in readings of the third law of thermodynamics or *Nernst’s unattainability principle* (UP), bounding achievable rates to cooling. Specifically, in this context, we consider the reading of the problem of preparing systems in a state which is arbitrarily close to being pure. Let us for simplicity take as an initial system two uncorrelated qubits  $\rho = \rho \otimes \rho$  with  $S(\rho) < 1/2$  (even the generalization to other systems is obvious). Theorem 1 then implies that it is possible to implement a transition



satisfying (1) and (2) so that the final state is  $\rho' = \rho' \otimes \mathbf{1}_2$ , where  $\mathbf{1}_k$  represents a maximally mixed state of dimension  $k$  and  $\rho'$  is any full-rank state with  $S(\rho') = \epsilon$  for arbitrarily small  $\epsilon > 0$ , i.e., arbitrarily close, in trace distance, to a pure state. This is reminiscent of protocols of *algorithmic cooling* [24–27] which take a large number  $n$  of “warm” qubits  $\rho$  and distill from them  $n_c = n[1 - S(\rho)]$  “cold” qubits having each a smallest eigenvalue  $\lambda_{\min} = \mathcal{O}(\exp(-n))$  (see, in particular, Ref. [24]). The advantage of our protocol is that we can obtain *arbitrarily cold systems* using a small number of copies,  $n = 2$  in this case, in contrast to the asymptotic i.i.d. setting considered in algorithmic cooling. Furthermore, the fact that the auxiliary systems remain invariant allows one to repeat the protocol for  $n/2$  copies of  $\rho$  using a single auxiliary system. Taking  $S(\rho) \approx 1/2$ , we obtain  $n_c \approx n/2$  qubits which are arbitrarily close to a pure state. This coincides with the bound given by algorithmic cooling, which in this case is  $n_c = n[1 - S(\rho)] \approx n/2$  and that is the ultimate bound for any entropy nondecreasing protocol. Hence, our protocol not only distills arbitrarily cold qubits with few copies, but also has an optimal efficiency—in terms of the rate of almost pure qubits—when applied sequentially in the asymptotic limit. At the same time, however, our protocol establishes correlations among the cold qubits produced. Hence, although they can be used individually for further applications, it would be wrong to conclude that using our results one can prepare an arbitrary number  $(\rho')^{\otimes n}$  of uncorrelated quasipure states using the same auxiliary system over and over (see Supplemental Material, Sec. III [20]). This again stresses the importance of correlations in the scheme.

The fact that one can produce systems in a state  $\rho'$  which is arbitrarily close to a pure state might, moreover, at first glance seem to be in contradiction with the third law of thermodynamics as formulated in the UP. The UP states that an infinite time is required to cool down a system to its ground state (see, e.g., Refs. [28–31] for recent approaches to quantum readings of the UP and their relation with pure state preparation). However, we note that preparing an arbitrarily pure  $\rho'$  also requires an arbitrarily large auxiliary system and might require a very large environment to implement the dephasing map  $\mathcal{D}$ , which, in turn, ensures that it cannot be prepared in a finite time.

*Relation to previous work.*—Let us now briefly discuss the relation of our results to previous work (see Fig. 2 for an overview). To begin with, we note that one can use previous results to fully characterize the possible state transitions  $\rho \rightarrow \rho'$  for the special case in which the auxiliary system is constrained to be a maximally mixed state. Specifically, one can recast recent results [32,33] as the statement that there exists a  $d$ -dimensional Hilbert space such that for any basis  $J$  of it there exists a unitary  $U$  such that

$$\text{Tr}_2[U(\rho \otimes \mathbf{1}_d)U^\dagger] = \rho', \quad (3)$$

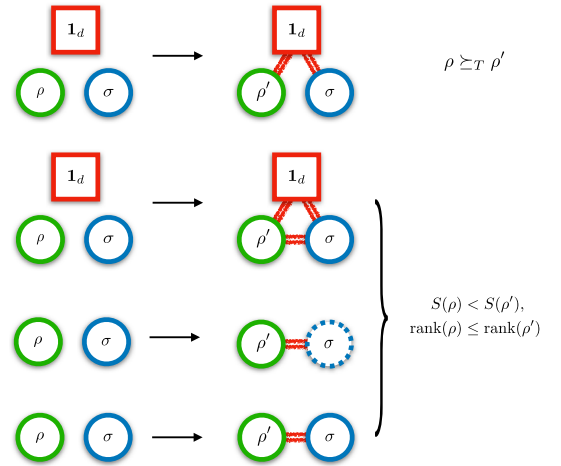


FIG. 2. Comparison of various settings and results. Top: State transitions implementable using a source of randomness and an uncorrelated catalyst  $\sigma$  are characterized by the trumping relations. Middle top: State transitions allowing for a source of randomness and a correlated catalyst, an auxiliary system that locally remains unchanged, are characterized by entropy and rank [18]. Middle bottom: By Theorem 1, state transitions using a correlated catalyst and a dephasing environment that acts on the catalyst (dashed boundary) are also characterized by entropy and rank. Bottom: State transitions using a correlated catalyst alone are characterized by entropy and rank. This is the content of Conjecture 1.

$$\mathcal{D}_J[\text{Tr}_1[U(\rho \otimes \mathbf{1}_d)U^\dagger]] = \mathbf{1}_d, \quad (4)$$

if and only if  $\rho$  majorizes  $\rho'$ , denoted by  $\rho \succeq \rho'$  [32]. Clearly, the above is a special case of Eqs. (1) and (2). Majorization captures the state transitions that are possible under random unitary evolution, and, hence, the above establishes the intuitive result that every random unitary evolution can be implemented with a sufficiently large source of randomness without affecting the latter’s state. To compare this result with Theorem 1, it should be noted that  $\rho \succeq \rho'$  is, as a constraint, much stronger than  $S(\rho') > S(\rho)$ . Indeed, one can see that Rényi entropies  $S_\alpha$ , defined as

$$S_\alpha(\rho) = \frac{1}{1-\alpha} \log \text{Tr}(\rho^\alpha), \quad (\alpha \in \mathbb{R} \setminus \{1\}), \quad (5)$$

cannot decrease for transitions  $\rho \rightarrow \rho'$  with  $\rho \succeq \rho'$ , where the VNE is given by the particular case of  $S \equiv S_1 := \lim_{\alpha \rightarrow 1} S_\alpha$ . The infinite set of conditions given by the Rényi entropies

$$S_\alpha(\rho') \geq S_\alpha(\rho) \quad \forall \alpha \in \mathbb{R} \quad (6)$$

become both necessary and sufficient for the existence of a further auxiliary system  $\sigma$  such that  $\rho \otimes \sigma \succeq \rho' \otimes \sigma$ —an important relation known as *trumping* [34,35] in quantum

information theory. The trumping constraints lie, in strength, strictly between those imposed by majorization and the VNE alone. Lastly, in Ref. [18], it is shown that by allowing for correlations between both systems it is possible to collapse the infinite set of conditions for the trumping conditions to essentially the VNE. In particular, it is shown that condition (i) in Theorem 1 is equivalent to the existence of  $\sigma$  and  $U$  so that  $\rho \otimes \sigma \succeq \rho'\sigma$ , where  $\rho'\sigma$  denotes a density matrix such that  $\text{Tr}_2(\rho'\sigma) = \rho'$  and  $\text{Tr}_1(\rho'\sigma) = \sigma$ . This statement differs from Theorem 1 in that one needs to make use of a maximally mixed system over which one has full unitary control, while Theorem 1 includes external randomness only in the form of an uncontrolled dephasing map (see Fig. 2 for a comparison).

*Catalytic entropy conjecture.*—The discussion above raises the natural question whether an external environment, being modeled as a maximally mixed state or a dephasing map as above, is at all necessary to implement all transitions which do not decrease the VNE. This is what we capture in the following conjecture.

*Conjecture 1: Catalytic entropy conjecture.*—Let  $\rho$  and  $\rho'$  be two density matrices of the same finite dimension and with different spectra. Then the following two statements are equivalent: (a)  $S(\rho') > S(\rho)$  and  $\text{rank}(\rho') \geq \text{rank}(\rho)$ . (b) There exists a density matrix  $\sigma$  and a unitary  $U$  such that

$$\text{Tr}_2[U(\rho \otimes \sigma)U^\dagger] = \rho' \quad \text{and} \quad \text{Tr}_1[U(\rho \otimes \sigma)U^\dagger] = \sigma. \quad (7)$$

The implication (b)  $\Rightarrow$  (a) follows directly from the subadditivity of the VNE and  $S_0$ ; hence, the real content of the conjecture is that (a) are the only constraints on transitions of the form (b). If true, this conjecture implies that the von Neumann entropy characterizes correlated catalytic state transitions in unitary quantum mechanics in full generality, without the need to introduce noise or i.i.d. limits (see Fig. 2).

Let us now discuss why we believe this conjecture to be true. To begin with, it is easy to generate counterexamples that rule out the possibility that transitions of the form (b) are constrained by the aforementioned trumping relations. In Supplemental Material [20], we provide such a counterexample together with a method to construct further examples. But, in fact, we can rule out more general constraints than (6) with the help of the following lemma.

*Lemma 2: Weak solution to catalytic entropy conjecture.*—Let  $\rho$  and  $\rho'$  be two density matrices of the same, finite dimension and with different spectra. Then the following two statements are equivalent: (I)  $S(\rho') > S(\rho)$  and  $\text{rank}(\rho') \geq \text{rank}(\rho)$ . (II) There exists a density matrix  $\sigma$ , a unitary  $U$ , and some finite dimension  $d$  such that

$$\text{Tr}_2[U(\rho \otimes \mathbf{1}_d \otimes \sigma)U^\dagger] = \rho' \otimes \mathbf{1}_d, \quad (8)$$

$$\text{Tr}_1[U(\rho \otimes \mathbf{1}_d \otimes \sigma)U^\dagger] = \sigma. \quad (9)$$

This result, which is proven in Supplemental Material [20], supports the conjecture in two ways: First, it shows that the catalytic entropy conjecture is true up to an additional maximally mixed system that remains uncorrelated to the system of interest but not to the auxiliary system. It can also be seen as an instance of the full catalytic entropy conjecture for the specific states  $\rho \otimes \mathbf{1}_d$  and  $\rho' \otimes \mathbf{1}_d$ . Second, and more importantly, it allows us to prove the following corollary.

*Corollary 3: Characterization of entropy functions.*—Let  $f$  be a function from the set of density matrices to the real numbers such that, for every transition of the form (b) between full-rank density matrices,  $f(\rho') > f(\rho)$ . Then exactly one of the following two statements is true: (1)  $S(\rho') > S(\rho) \Leftrightarrow f(\rho') > f(\rho)$ , (2)  $f$  is nonadditive or discontinuous.

Corollary 3 follows from Lemma 2 by showing that any such function  $f$  has to be a linear function of the VNE (see Supplemental Material, Sec. V [20]). Thus, for full-rank density matrices, if Conjecture 1 was false, any additional constraint on transitions of the form (b) would have to be given by exotic entropic functions that are not additive or are discontinuous. For instance, this corollary immediately implies that none of the functions  $S_\alpha$ ,  $\alpha \neq 0, 1$ , can be a monotone for transitions of the form (b), since they all satisfy none of the two conditions in the corollary.

*Discussion and open questions.*—In this Letter, we have provided a new operational characterization of von Neumann entropy which adds significant support to recent proposals that, contrary to common wisdom, the standard von Neumann entropy characterizes not only the i.i.d. limit but also single-shot protocols in quantum information theory. We have done so by showing that the von Neumann entropy fully determines the possibility of single-shot state transitions in unitary quantum mechanics, as long as one has access to a catalyst, which may build up correlations, and environmental dephasing in a preferred basis. Furthermore, we have formulated the catalytic entropy conjecture which essentially states that the above result holds true even in the absence of decoherence. We have also presented evidence for the truth of this conjecture by ruling out alternatives.

Our work suggests that there might be a novel, hitherto unexplored sector of quantum information theory in which operations on *single* copies of a quantum state are characterized directly in terms of standard entropic quantities like VNE. For example, one may ask what happens in Theorem 1 or Conjecture 1 if we introduce another reference system  $R$  that is initially correlated or entangled with the system 1 (let us denote system 1 by  $A$  for now, and let  $C$  be the catalytic auxiliary system 2). Applying a unitary  $U_{A,C}$  on  $A$  and  $C$ , denoting the new states of the systems by  $R'$ ,  $A'$ , and  $C'$ , we obtain  $R' = R$ , by construction  $C' = C$ , and  $S(A') \geq S(A)$ , since  $A$  becomes correlated with  $C$ . Furthermore, the mutual information



$I(R:A)=S(R)+S(A)-S(R,A)$  satisfies  $I(R':A')\leq I(R:A)$ . Are these necessary conditions also *sufficient* for the existence of a transformation of that form—in particular, can  $A$  retain almost all of its correlations with  $R$  under correlating-catalytic transformations? A positive answer to this or other similar questions would yield a new single-shot interpretation of the standard mutual information which could potentially be useful in the context of *decoupling* [5,36–38] or merging of quantum states.

The results also hint at the insight that entanglement in single many-body systems can well be captured in terms of the von Neumann entropy. Ideas on *single-copy entanglement* have been considered in situations where each specimen consists of a many-body system, already naturally featuring asymptotically many constituents [39]. Then it can be unreasonable to capture entanglement of subsystems in yet another asymptotic limit of many copies of identical quantum many-body systems. The results laid out here give substance to the intuition that, even in single specimens of quantum many-body systems, entanglement can in this context be quantified in terms of the familiar von Neumann entanglement entropy.

Results of this kind would also have implications in the context of *holographic approaches* to quantum gravity, as in the AdS/CFT correspondence (see, for example, Refs. [40–47]). In these approaches, standard von Neumann (entanglement) entropies of boundary regions turn out to correspond to geometric quantities of a dual gravity theory in the bulk. In fact, it is exactly the mutual information that we have just discussed which is believed to be directly related to geometric quantities like area also in other (non-AdS/CFT) approaches to emergent spacetime [48–50]. To shed some light on this correspondence, it is therefore natural to consider operational interpretations of entropy in the boundary theory and to “dualize” them to obtain corresponding interpretations of geometric quantities in the bulk. A difficulty in doing so, however, is that the protocols on the boundary theory either involve many copies of the state (which seems unphysical given that there is a unique spacetime) or lead to quantification in terms of single-shot entropies (see, e.g., Ref. [44]) which do not always have a direct dual interpretation. The proven and conjectured results of this letter could therefore resolve this difficulty, by supplying a direct single-shot interpretation of standard entropic quantities which might ultimately shed some light on the operational basis of geometric quantities.

We acknowledge funding from Deutsche Forschungsgemeinschaft (GA 2184/2-1, CRC 183, EI 519/14-1, EI 519/9-1, FOR 2724), the European Research Council (TAQ), and the Studienstiftung des deutschen Volkes. H. W. further acknowledges contributions from the Swiss National Science Foundation via the NCCR QSIT as well as Project No. 200020\_165843. This research was supported in part by Perimeter Institute for

Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science.

- [1] B. Schumacher, *Phys. Rev. A* **51**, 2738 (1995).
- [2] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, *Phys. Rev. A* **53**, 2046 (1996).
- [3] F. Hiai and D. Petz, *Commun. Math. Phys.* **143**, 99 (1991).
- [4] T. Ogawa and H. Nagaoka, in *Asymptotic Theory of Quantum Statistical Inference* (World Scientific, Singapore, 2005), pp. 28–42.
- [5] C. Majenz, M. Berta, F. Dupuis, R. Renner, and M. Christandl, *Phys. Rev. Lett.* **118**, 080503 (2017).
- [6] M. Mosonyi and T. Ogawa, *Commun. Math. Phys.* **334**, 1617 (2015).
- [7] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, *Phys. Rev. Lett.* **111**, 250404 (2013).
- [8] N. Datta, *IEEE Trans. Inf. Theory* **55**, 2816 (2009).
- [9] M. Berta, K. P. Seshadreesan, and M. M. Wilde, *J. Math. Phys. (N.Y.)* **56**, 022205 (2015).
- [10] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, *J. Math. Phys. (N.Y.)* **54**, 122203 (2013).
- [11] M. M. Wilde, A. Winter, and D. Yang, *Commun. Math. Phys.* **331**, 593 (2014).
- [12] R. Renner, Ph.D. thesis, ETH Zurich, 2005.
- [13] N. D. R. Renner, *IEEE Trans. Inf. Theory* **55**, 2807 (2009).
- [14] M. P. Müller and M. Pastena, *IEEE Trans. Inf. Theory* **62**, 1711 (2016).
- [15] M. Lostaglio, M. P. Müller, and M. Pastena, *Phys. Rev. Lett.* **115**, 150402 (2015).
- [16] R. Gallego, J. Eisert, and H. Wilming, *New J. Phys.* **18**, 103017 (2016).
- [17] H. Wilming, R. Gallego, and J. Eisert, *Entropy* **19**, 241 (2017).
- [18] M. P. Müller, *Phys. Rev. X* **8**, 041051 (2018).
- [19] A. M. Alhambra, L. Masanes, J. Oppenheim, and C. Perry, [arXiv:1709.06139](https://arxiv.org/abs/1709.06139).
- [20] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.122.210402> for the proof, which includes Refs. [21–23], in Sec. I, a discussion on the protocol applied sequentially and the correlations that it establishes between subsystems in Sec. III, and, in particular, a proof of Corollary 3 in Sec. V.
- [21] A. Horn, *Am. J. Math.* **76**, 620 (1954).
- [22] J. Schwinger, *Proc. Natl. Acad. Sci. U.S.A.* **46**, 570 (1960).
- [23] R. F. Werner, *J. Phys. A* **34**, 7081 (2001).
- [24] L. J. Schulman and U. V. Vazirani, in *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing—STOC '99* (ACM, New York, 1999), p. 322.
- [25] L. J. Schulman, T. Mor, and Y. Weinstein, *Phys. Rev. Lett.* **94**, 120501 (2005).
- [26] P. O. Boykin, T. Mor, V. Roychowdhury, F. Vatan, and R. Vrijen, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3388 (2002).
- [27] S. Raeisi and M. Mosca, *Phys. Rev. Lett.* **114**, 100404 (2015).

- [28] A. Levy, R. Alicki, and R. Kosloff, *Phys. Rev. E* **85**, 061126 (2012).
- [29] J. Scharlau and M. P. Müller, *Quantum* **2**, 54 (2018).
- [30] L. Masanes and J. Oppenheim, *Nat. Commun.* **8**, 14538 (2017).
- [31] H. Wilming and R. Gallego, *Phys. Rev. X* **7**, 041033 (2017).
- [32] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Yunger Halpern, *Phys. Rep.* **583**, 1 (2015).
- [33] P. Boes, H. Wilming, R. Gallego, and J. Eisert, *Phys. Rev. X* **8**, 041016 (2018).
- [34] M. Klimesh, arXiv:0709.3680v1.
- [35] S. Turgut, *J. Phys. A* **40**, 12185 (2007).
- [36] M. Horodecki, J. Oppenheim, and A. Winter, *Nature (London)* **436**, 673 (2005).
- [37] P. Hayden, Tutorial QIP 2011, Singapore, 2011 (unpublished), <https://qip2011.quantumlah.org/tutorialprogramme/>.
- [38] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner, *Commun. Math. Phys.* **328**, 251 (2014).
- [39] J. Eisert and M. Cramer, *Phys. Rev. A* **72**, 042112 (2005).
- [40] L. Susskind, *J. Math. Phys. (N.Y.)* **36**, 6377 (1995).
- [41] J. M. Maldacena, *Int. J. Theor. Phys.* **38**, 1113 (1999).
- [42] S. Ryu and T. Takayanagi, *Phys. Rev. Lett.* **96**, 181602 (2006).
- [43] P. Hayden, M. Headrick, and A. Maloney, *Phys. Rev. D* **87**, 046003 (2013).
- [44] B. Czech, P. Hayden, N. Lashkari, and B. Swingle, *J. High Energy Phys.* **06** (2015) 157.
- [45] N. Lashkari and M. Van Raamsdonk, *J. High Energy Phys.* **04** (2016) 153.
- [46] H. Casini, E. Testé, and G. Torroba, *Phys. Rev. Lett.* **118**, 261602 (2017).
- [47] A. Jahn, M. Gluza, F. Pastawski, and J. Eisert, arXiv:1711.03109.
- [48] M. V. Raamsdonk, *Gen. Relativ. Gravit.* **42**, 2323 (2010).
- [49] C. J. Cao, S. M. Carroll, and S. Michalakis, *Phys. Rev. D* **95**, 024031 (2017).
- [50] C. J. Cao and S. M. Carroll, *Phys. Rev. D* **97**, 086003 (2018).

## Supplemental material: Von Neumann entropy from unitarity

Paul Boes,<sup>1</sup> Jens Eisert,<sup>1</sup> Rodrigo Gallego,<sup>1</sup> Markus P. Müller,<sup>2,3</sup> and Henrik Wilming<sup>1,4</sup>

<sup>1</sup>Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany

<sup>2</sup>Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Boltzmannngasse 3, A-1090 Vienna, Austria

<sup>3</sup>Perimeter Institute for Theoretical Physics, Waterloo, ON N2L 2Y5, Canada

<sup>4</sup>Institute for Theoretical Physics, ETH Zurich, 8093 Zurich, Switzerland

### I. PROOF OF THEOREM 1 AND LEMMA 2

In this section we prove Theorem 1 and Lemma 2 in the main text. The proofs of both results rely on the following recent result from Ref. [1].

**Theorem 6** (Correlating-catalytic majorization [1]). *Let  $\rho, \rho'$  be two density matrices on the same, finite-dimensional Hilbert space  $\mathcal{H}_A$  such that  $S(\rho) < S(\rho')$  and  $\text{rank}(\rho) \leq \text{rank}(\rho')$ . Then there exists a density matrix  $\tau$  on a finite-dimensional Hilbert space  $\mathcal{H}_B$  and a bipartite density matrix  $\rho'\tau$  on  $\mathcal{H}_A \otimes \mathcal{H}_B$  such that*

$$\rho \otimes \tau \succeq \rho'\tau, \quad \text{Tr}_B[\rho'\tau] = \rho', \quad \text{Tr}_A[\rho'\tau] = \tau.$$

Another result that will be used frequently is the Schur-Horn-Theorem.

**Theorem 7** (Schur-Horn [2]). *For a matrix  $H$ , let  $\lambda(H)$  be the vector of its eigenvalues and  $\text{diag}(H)$  the vector of its diagonal entries. If  $H$  is Hermitian, then the following are equivalent:*

- $\lambda(H) \succeq \text{diag}(H)$ ,
- there exists a unitary matrix  $U$  such that

$$U\hat{\lambda}(H)U^\dagger = H,$$

where  $\hat{\lambda}(H)$  is the diagonal matrix with diagonal  $\lambda(H)$ .

In particular, the Schur-Horn theorem implies that, if  $\rho \succeq \rho'$ , then there exist unitaries  $U, V$  such that

$$\rho' = V(\mathcal{D}_J[U\rho U^\dagger])V^\dagger. \quad (1)$$

Here and in the following, in contrast to the main text, we explicitly denote the choice of basis  $J = \{|j\rangle\}$  in the notation for the decoherence map,  $\mathcal{D} = \mathcal{D}_J$ . If we choose  $J$  as the eigenbasis of  $\rho'$  then  $V$  is the identity map. We are now in position to prove Theorem 1 in the main text.

*Proof of Theorem 1.* The proof of Theorem 1 proceeds in several steps. To aid understanding, Fig. 1 provides an overview over the various steps.

We begin with proving that i) implies ii). Thus, assume that  $S(\rho) < S(\rho')$  and  $\text{rank}(\rho) \leq \text{rank}(\rho')$ . Then Theorem 6 together with (1) implies that there exists a  $d_B$ -dimensional Hilbert space, a state  $\tau$  on this space, a unitary  $W_{A,B}$  and two bases  $J_A$  and  $J_B$  such that

$$(\mathcal{D}_{J_A} \otimes \mathcal{D}_{J_B}) [W_{A,B}(\rho \otimes \tau)W_{A,B}^\dagger] = \rho'\tau.$$

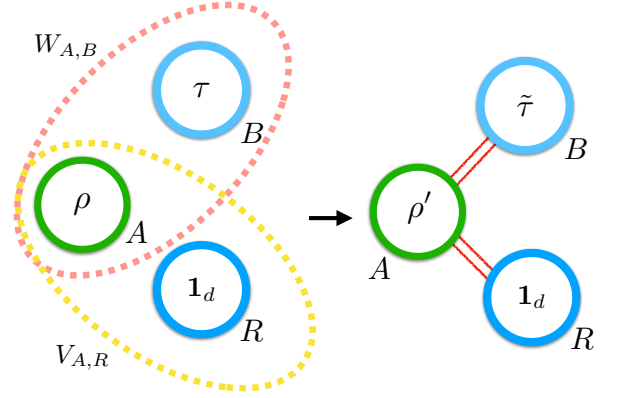


FIG. 1. Proof Sketch for Theorem 1. We consider three systems  $A, B, R$ .  $A$  corresponds to system 1 in the main text, while systems  $B$  and  $R$  jointly correspond to system 2 in the main text. We then initiate this tripartite system in the state  $\rho_A \otimes \tau_B \otimes (\mathbf{1}_d)_R$ , where  $\tau$  is a state that depends on both the initial and final states on  $A$  and where  $\mathbf{1}_d$  is the maximally mixed state with  $d$  being the dimension of  $A$ . To this tripartite state we apply two unitaries  $W_{A,B}, V_{A,R}$  (represented by dashed ellipses), one with support only on  $A, B$ , the other with support only  $A, R$ . These two unitaries have the effect of producing a global state with reduced density matrices  $\rho'_A, \tilde{\tau}_B$ , and  $(\mathbf{1}_d)_R$  and in which, importantly, the reduced state on  $B$  and  $R$  is product. Here,  $\tilde{\tau}$  is a state with the property that  $\mathcal{D}_J(\tilde{\tau}) = \tau$ . Hence, after dephasing system  $BR$  (and hence  $B$ ),  $BR$  is returned back to its initial state. The existence of  $\tau, W_{A,B}, V_{A,R}$  for any pair of states  $\rho, \rho'$  to which the Theorem applies then establishes the claim.

Here,  $J_B$  is the eigenbasis of  $\tau$  and hence can be chosen at will by adjusting the unitary  $W_{A,B}$  and the state  $\tau$ . From locality of quantum mechanics and the Schur-Horn theorem we thus find that

$$\rho'\tilde{\tau} := (\mathcal{D}_{J_A} \otimes \mathbb{I}) [W_{A,B}(\rho \otimes \tau)W_{A,B}^\dagger]$$

is a quantum state with the properties  $\text{Tr}_B[\rho'\tilde{\tau}] = \rho'$  and  $\tilde{\tau} = \text{Tr}_A[\rho'\tilde{\tau}] \succeq \tau$ . Here,  $\mathbb{I}$  denotes the identity super-operator.

As a second step, we show that we can realize any dephasing map on a system  $A$  using an ancillary system in a maximally mixed state. To see this, let  $R$  be a system of the same dimension  $d$  as  $A$  and let  $\{U_k\}_{k=1}^d$  be a unitary operator basis on  $A$ , meaning a collection of  $d$  unitaries  $U_k$  such that

$$\text{Tr} [U_j U_k^\dagger] = d\delta_{j,k}. \quad (2)$$

Such a set of operators exists on every finite-dimensional Hilbert space [3, 4]. Then, define the unitary

$$V_{A,R} = \sum_{j=1}^d |j\rangle\langle j|_A \otimes (U_j)_R,$$

where we recall that  $J = \{|j\rangle\}$ . Now, it is easy to check that for any  $\rho = \rho_A$ ,

$$\text{Tr}_R [V_{A,R}(\rho \otimes \mathbf{1}_d)V_{A,R}^\dagger] = \mathcal{D}_J[\rho].$$

In a third step, we now show that we can use this dilation of the dephasing map to construct an auxiliary system for Theorem 1 in the main text. To do so, let

$$\sigma := \tau \otimes \mathbf{1}_d$$

and define the unitary

$$U_{A,B,R} = (V_{A,R} \otimes \mathbf{1}_B)(W_{A,B} \otimes \mathbf{1}_R).$$

We emphasize at this point that the auxiliary system has dimension  $d_B d$  and for any choice of basis on the Hilbert space of the latter we can find a tensor factorization as above and a corresponding quasi-classical state  $\tau$  (the maximally mixed state is of course quasi-classical in any basis) and corresponding unitaries  $W_{A,B}$  and  $V_{A,B}$ . From the previous discussion and the construction of the dephasing unitary  $V_{A,R}$ , we know that

$$\text{Tr}_R [U_{A,B,R}(\rho \otimes \sigma)U_{A,B,R}^\dagger] = \rho' \tilde{\tau}.$$

Thus, what is left to be proven is that  $\sigma$  is valid, i.e., does not change in the course of the process except from building up coherences. We will show that it undergoes the transition

$$\sigma = \tau \otimes \mathbf{1}_d \rightarrow \tilde{\tau} \otimes \mathbf{1}_d.$$

To show this, first note that the dephasing dilation implemented by  $V_{A,R}$  leaves the state  $\mathbf{1}_d$  of  $R$  locally unchanged. But this means that we only have to show that  $R$  does not become correlated with  $B$  in the dephasing step, since it follows from locality that the marginal on  $R$  remains unchanged and the marginal on  $B$  evolves from  $\tau$  to  $\tilde{\tau}$ . To see that  $B$  and  $R$  remain uncorrelated, we simply compute the action of the dephasing unitary  $V_{A,R}$  on  $B, R$ , to get

$$\begin{aligned} & \text{Tr}_A [U_{A,B,R}(\rho \otimes \sigma)U_{A,B,R}^\dagger] \\ &= \sum_{j,k} \text{Tr}_A [ |j\rangle\langle j|_A W_{A,B}(\rho \otimes \tau)W_{A,B}^\dagger |k\rangle\langle k|_A ] \otimes \frac{U_j U_k^\dagger}{d_R} \\ &= \sum_j \langle j|_A W_{A,B}(\rho \otimes \tau)W_{A,B}^\dagger |j\rangle_A \otimes \mathbf{1}_d \\ &= \tilde{\tau} \otimes \mathbf{1}_d, \end{aligned}$$

where we have dropped identities for notational convenience. This proves that i) implies ii).

$$\begin{array}{ccc|c} [\gamma_{A,B}]_{0,0} & [\gamma_{A,B}]_{0,1} & [\gamma_A]_0 & 0 \\ [\gamma_{A,B}]_{1,0} & [\gamma_{A,B}]_{1,1} & [\gamma_A]_1 & 0 \\ [\gamma_{A,B}]_{2,0} & [\gamma_{A,B}]_{2,1} & [\gamma_A]_2 & 0 \\ \hline [\gamma_B]_0 & [\gamma_B]_1 & & \end{array} ; \quad \begin{array}{c} 0 \\ 2 \\ 2 \\ 2 \\ \hline 2 \\ 2 \\ 2 \\ 2 \\ \hline 2 \\ 2 \\ 2 \\ 2 \end{array} \begin{array}{c} 0 \\ 1 \\ 1 \\ 1 \\ \hline 1 \\ 1 \\ 1 \\ 1 \\ \hline 1 \\ 1 \\ 1 \\ 1 \end{array} \begin{array}{c} 0 \\ 1 \\ 1 \\ 1 \\ \hline 1 \\ 1 \\ 1 \\ 1 \\ \hline 1 \\ 1 \\ 1 \\ 1 \end{array} \begin{array}{c} 0 \\ 1 \\ 1 \\ 1 \\ \hline 1 \\ 1 \\ 1 \\ 1 \\ \hline 1 \\ 1 \\ 1 \\ 1 \end{array} \rightarrow \begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 2 & 2 \\ 1 & 0 & 2 & 2 \\ \hline 1 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 \\ \hline 1 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 \end{array}$$

FIG. 2. Counterexample to show that trumping relations cannot characterize transitions of the form (b) of Conjecture 1 in the main text (for notation see Appendix II below)

Let us now prove that ii) implies i). In the following let  $\alpha \in \{0, 1\}$ . Since  $S_0(\rho) = \log(\text{rank}(\rho))$ , both  $S_0$  and  $S_1 = S$  are subadditive and additive. Since the final state on the auxiliary system, which we now call  $\sigma'$ , satisfies  $\mathcal{D}_J[\sigma'] = \sigma$ , it follows that  $\sigma' \succeq \sigma$  and thus  $S_\alpha(\sigma') \leq S_\alpha(\sigma)$ . Furthermore, from additivity and subadditivity we get

$$\begin{aligned} S_\alpha(\rho) + S_\alpha(\sigma) &= S_\alpha(\rho \otimes \sigma) = S_\alpha(\rho' \sigma') \\ &\leq S_\alpha(\rho') + S_\alpha(\sigma') \leq S_\alpha(\rho') + S_\alpha(\sigma). \end{aligned}$$

For  $\alpha = 0$ , this proves  $\text{rank}(\rho) \leq \text{rank}(\rho')$ . For  $\alpha = 1$ , equality, i.e.  $S(\rho) = S(\rho')$ , is only possible if  $S(\rho' \sigma') = S(\rho') + S(\sigma')$ , and it is well-known that this implies  $\rho' \sigma' = \rho' \otimes \sigma'$ . Thus  $\rho \otimes \sigma \succ \rho' \otimes \sigma' \succ \rho' \otimes \sigma$ , and so  $\rho \succ_T \rho'$  for the trumping relation, which together with  $S(\rho) = S(\rho')$  implies that  $\rho$  and  $\rho'$  have the same spectrum, i.e. are unitarily equivalent [5, 6]. This contradicts the assumptions of the theorem. We must thus have  $S(\rho) < S(\rho')$ , which completes the proof.  $\square$

Let us now turn to the proof of Lemma 2 in the main text, which builds on the proof of Theorem 1.

*Proof of Lemma 2 in the main text.* For this proof we re-use all the notation from the proof of the implication i) $\Rightarrow$ ii) in the proof given above. In particular note that the final state  $\tilde{\tau}$  on the  $B$ -subsystem of the auxiliary system only needs to be dephased in a basis  $J_B$  to be returned exactly, since, by construction,  $\text{diag}(\tilde{\tau}) = \lambda(\tau)$ . Using the dephasing construction already used in the proof of Theorem 1 we can include a further system  $R_2$  in the maximally mixed state into the system and use the dephasing unitary  $V_{R_2 B}$  at the end of the process to dephase system  $B$ . The only property that we still need to prove is that this does not introduce correlations between  $A$  and  $R_2$ . However, this is exactly the same calculation that shows that there are no correlations between  $B$  and  $R$  at the end in the proof of Theorem 1. We only have to exchange  $R$  for  $R_2$  and  $B$  for  $A$ . This finishes the proof.  $\square$

## II. TRUMPING RELATIONS CANNOT CHARACTERIZE CATALYTIC TRANSITIONS

Here, we show that the trumping relations, Eq. (6) in the main text, cannot characterize transitions of the form (b) in the main text, by means of a counterexample. In order to understand this construction, it is helpful to consider the table in Fig. 2: Given an arbitrary bipartite state on  $A, B$  denoted  $\gamma_{A,B}$ , the table at the left-hand side indicates the meaning of each entry, where  $[\gamma_{A,B}]_{i,j} := \langle i, j | \gamma_{A,B} | i, j \rangle$  on

a given computational basis of  $AB$ . The two tables at the right hand side indicate a particular transition of the form  $\rho \otimes \sigma \rightarrow U(\rho \otimes \sigma)U := \rho'\sigma$ . In this case we take  $\rho$  and  $\sigma$  to be of dimension 3 and 2 respectively, and both diagonal in the computational basis. The unitary  $U$  is simply a classical permutation which swaps the red entries with the blue entries. Note that the final state satisfies  $\text{Tr}_2(\rho'\sigma) = \sigma$  since the bottom row remains unchanged, as demanded by condition (b). The row sums on the right-hand side of each table represent  $\rho = \text{diag}(0, 1/2, 1/2)$  and  $\rho' = \text{diag}(1/6, 1/6, 2/3)$ . Since  $S_\infty(\rho)$  is determined by the largest eigenvalue of  $\rho$ , this example realizes a catalytic transition  $\rho \rightarrow \rho'$  with  $S_\infty(\rho) > S_\infty(\rho')$  and hence excludes the possibility that catalytic state transitions are constrained by the trumping relations.

### III. CATALYTIC COOLING

Let us first present in detail how to prepare almost pure states with a protocol that uses Theorem 1. Using this theorem, we have that, given system  $Q_1$  in state  $\rho_{Q_1} = \varrho \otimes \varrho$  with  $2S(\varrho) < 1$ , one can find  $U$  and an auxiliary system  $C$  in state  $\sigma$  so that

$$\gamma_{Q_1 C} = (\mathcal{D}_J \circ \mathcal{U}_1)[\rho_{Q_1} \otimes \sigma_C] \quad (3)$$

where  $\mathcal{D}_J$  is the map locally dephasing the system  $C$  and leaving  $Q_1$  untouched (formally  $\mathbb{I}_{Q_1} \otimes \mathcal{D}_J$ ), and  $\mathcal{U}_1[\bullet] = U \bullet U^\dagger$ . Also, we denote by  $\gamma_{Q_1 C}$  a bipartite state on  $Q_1 C$  which, according to Theorem 1, fulfills  $\text{Tr}_{Q_1}(\gamma_{Q_1 C}) = \sigma_C$  and  $\text{Tr}_C(\gamma_{Q_1 C}) = \rho'_{Q_1} = \varrho' \otimes \mathbf{1}_2$ , where  $\varrho'$  can be any full-rank state, but in the following we are interested in the case where  $\rho'$  is arbitrarily close to a pure state.

This protocol can be iterated on an arbitrary number  $n$  of subsystems  $Q_1, \dots, Q_n$ , taking initially  $\rho_{Q_1, \dots, Q_n} = \rho_{Q_1} \otimes \dots \otimes \rho_{Q_n}$  as input, where  $\rho_{Q_i} = \varrho \otimes \varrho$  for all  $i$ . We define the unitary channels  $\mathcal{U}_i$  which apply the unitary  $U$  to systems  $Q_i C$  and act trivially in the rest of the subsystems, that is,

$$\mathcal{U}_i[\bullet] = U_{Q_i C} \otimes \mathbb{I}_{|Q_i C} \bullet U_{Q_i C}^\dagger \otimes \mathbb{I}_{|Q_i C}. \quad (4)$$

Then, applying these unitary channels, each followed by a dephasing map on  $C$ , one obtains

$$\gamma_{Q_1, \dots, Q_n C} = \mathcal{D}_J \circ \mathcal{U}_n \circ \dots \circ \mathcal{D}_J \circ \mathcal{U}_1[\rho_{Q_1, \dots, Q_n} \otimes \sigma] \quad (5)$$

where, due to Theorem 1, we have

$$\begin{aligned} \text{Tr}_{|Q_i}(\gamma_{Q_1, \dots, Q_n C}) &= \varrho' \otimes \mathbf{1}_2 \quad \forall i, \\ \text{Tr}_{|C}(\gamma_{Q_1, \dots, Q_n C}) &= \sigma. \end{aligned}$$

Hence, with this protocol we have prepared  $n/2$  subsystems whose marginal  $\varrho'$  is arbitrarily close to a pure state. Note, however, that the resulting state of the compound  $\gamma_{Q_1, \dots, Q_n}$  displays correlations between its parts, hence, although each subsystem in state  $\varrho'$  can be individually used—for instance as a pure state input of a quantum computation—the whole compound  $\gamma_{Q_1, \dots, Q_n}$  deviates from the state

$$\tilde{\gamma}_{Q_1, \dots, Q_n} := \rho'_{Q_1} \otimes \dots \otimes \rho'_{Q_n}.$$

This can be seen for instance by comparing the minimum eigenvalue  $\lambda_{\min}$  of both states in the limit of large  $n$ , which gives

$$\lim_{n \rightarrow \infty} \frac{\lambda_{\min}(\tilde{\gamma}_{Q_1, \dots, Q_n})}{\lambda_{\min}(\gamma_{Q_1, \dots, Q_n})} \leq \lim_{n \rightarrow \infty} \frac{\lambda_{\min}(\tilde{\gamma}_{Q_1, \dots, Q_n})}{\lambda_{\min}(\gamma_{Q_1, \dots, Q_n})} \quad (6)$$

$$\leq \lim_{n \rightarrow \infty} \frac{\lambda_{\min}(\tilde{\gamma}_{Q_1, \dots, Q_n})}{\lambda_{\min}(\rho_{Q_1, \dots, Q_n} \otimes \sigma)} \quad (7)$$

$$= \lim_{n \rightarrow \infty} \frac{(\frac{1}{2}\lambda_{\min}(\varrho'))^n}{\lambda_{\min}(\varrho)^{2n} \lambda_{\min}(\sigma)} \quad (8)$$

$$= 0 \quad (9)$$

where (6) follows simply because tracing out one subsystem can only increase the minimum eigenvalue; (7) follows due to (5). To see this note that map  $\mathcal{D}_J \circ \mathcal{U}_n \circ \dots \circ \mathcal{D}_J$  can be implemented as a global unitary on  $Q_1, \dots, Q_n$  together with a source of randomness of sufficiently large dimension  $d$  which is responsible of the dephasing. That is, there exists  $V$  so that

$$\text{Tr}(V \rho_{Q_1, \dots, Q_n} \otimes \sigma \otimes \mathbf{1}_d V^\dagger) = \gamma_{Q_1, \dots, Q_n C}.$$

This implies in turn that  $\rho_{Q_1, \dots, Q_n} \otimes \sigma \succeq \gamma_{Q_1, \dots, Q_n C}$  (see for instance Ref. [7]) and that

$$\lambda_{\min}(\rho_{Q_1, \dots, Q_n} \otimes \sigma) \leq \lambda_{\min}(\gamma_{Q_1, \dots, Q_n C}).$$

Eq. (8) follows from simple algebra. Lastly, (9) follows from the fact that  $\lambda_{\min}(\varrho')$  is arbitrarily small while  $\lambda_{\min}(\sigma) > 0$ . To see the latter we recall the result of Appendix F.1 from Ref. [8], which shows that any transition of the form (5) employing an auxiliary system  $\sigma$  without full rank with spectrum  $\{\sigma_i\}$ , can be also implemented with a full-rank state  $\tilde{\sigma}$  with spectrum  $\{\sigma_i | \sigma_i > 0\}$ . In other words, we can assume without loss of generality that  $\sigma$  is full rank.

### IV. THE CLASSICAL CASE

In the following, we denote the marginals of a probability distribution  $r$  on  $X \times Y$  by  $r_X$  resp.  $r_Y$ , such that  $r_X(x) = \sum_{y \in Y} r(x, y)$  and  $r_Y(y) = \sum_{x \in X} r(x, y)$ . This is the classical analogue of the partial trace.

**Conjecture 1** (Classical catalytic entropy conjecture). *Let  $p$  and  $p'$  be two different probability distributions on a finite space of events  $X$ . Then the following two statements are equivalent:*

- i)  $S(p) \leq S(p')$ , where  $S$  is the Shannon entropy.
- ii) For every  $\epsilon > 0$ , there exists a probability distribution  $q$  on a finite space  $Y$  and a permutation  $P$  on  $X \times Y$  such that

$$[P(p \otimes q)]_Y = q, \quad \|[P(p \otimes q)]_X - p'\|_1 \leq \epsilon. \quad (10)$$

There are two reasons for which we only conjecture approximability of  $p'$  to arbitrary accuracy instead of perfect achievability. Firstly, in order to drop the rank condition from condition i); secondly, to account for the case in which  $p$  and  $p'$

differ by irrational amounts. In this case, permutations only realize the transition  $p \rightarrow p'$  approximately.

Note that, since the statement of the catalytic entropy conjecture is unitarily invariant on the input states, and permutations are special cases of unitary operations, a proof of the classical catalytic entropy conjecture would essentially also prove the quantum version. The converse, however, is not necessarily true: it is a priori possible that only the quantum formulation holds. Nevertheless, as in the quantum case, one can show that the Shannon entropy is essentially the unique additive monotone. This follows from the following classical version of Lemma 2 in the main text. It uses the notation  $\text{rank}(p)$  to denote the number of non-zero entries of a discrete probability distribution  $p$ .

**Lemma 8** (Weak solution to catalytic entropy conjecture (classical)). *Let  $p$  and  $p'$  be two different probability distributions of the same, finite dimension and with rational entries. Then the following two statements are equivalent:*

- (I)  $S(p') > S(p)$  and  $\text{rank}(p') \geq \text{rank}(p)$ .
- (II) *There exists a probability distribution  $q$  on a finite sample space  $Z$ , a  $d$ -dimensional sample space  $Y$ , and a permutation  $P$  on  $X \times Y \times Z$  such that*

$$[P(p \otimes \mathbf{1}_d \otimes q)]_Z = q, \quad [P(p \otimes \mathbf{1}_d \otimes q)]_{X,Y} = p' \otimes \mathbf{1}_d.$$

Here,  $\mathbf{1}_d^\top = (1/d, \dots, 1/d)$  denotes the uniform distribution on  $Y$ .

*Proof.* We only consider the non-obvious direction, i.e. we show that (I) $\Rightarrow$ (II). According to Ref. [1], if condition (I) is satisfied, then there exists a probability distribution  $\tilde{q}$  on some sample space  $\tilde{Z}$  such that  $p \otimes \tilde{q} \succeq p' \tilde{q}$ , where  $p' \tilde{q}$  denotes a probability distribution on  $X \times Y$  such that  $[p' \tilde{q}]_Z = \tilde{q}$  and  $[p' \tilde{q}]_X = p'$ . Since  $p, p'$  are rational, and so are  $\tilde{q}$  and  $p' \tilde{q}$ , the majorization relation implies that this transition can be realized exactly with a random permutation. In other words, there exists a  $\tilde{d}$ -dimensional ancilla  $A$  in the state  $\mathbf{1}_{\tilde{d}}$  and the global permutation  $P = \sum_{i=1}^{\tilde{d}} \Pi_i \otimes P_i$ , where  $\Pi_i$  denotes the rank-one projector onto the standard basis  $\{\mathbf{e}_i\}$  of  $A$ , that is,  $\Pi_i(q) = q_i \mathbf{e}_i$ , such that

$$\frac{1}{\tilde{d}} \sum_{i=1}^{\tilde{d}} P_i(p \otimes \tilde{q}) = p' \tilde{q}. \quad (11)$$

Next, choose  $d = \tilde{d}$  as the dimension of  $Y$  and consider the permutation

$$P' = \sum_{i=1}^d (\Pi_i)_Y \otimes (\pi^i)_A, \quad (12)$$

where  $\pi$  is a permutation defined by  $\pi \mathbf{e}_j = \mathbf{e}_{j+1 \bmod d}$ . Ap-

plying both of these permutations to the total system yields

$$\begin{aligned} & P' P [(\mathbf{1}_{\tilde{d}})_Y \otimes (\mathbf{1}_{\tilde{d}})_A \otimes p \otimes \tilde{q}] \\ &= P' \left[ (\mathbf{1}_{\tilde{d}})_Y \otimes \left( \sum_i^d (\mathbf{e}_i)_A / d \otimes P_i(p \otimes \tilde{q}) \right) \right] \\ &= \sum_{i,j=1}^d (\mathbf{e}_j)_Y / d \otimes (\mathbf{e}_{i+j \bmod d})_A / d \otimes P_i(p \otimes \tilde{q}). \end{aligned}$$

From the last expression, we see that summing over  $Y$  leaves  $A$  uncorrelated from both  $X$  and  $\tilde{Z}$ , since  $\sum_j \Pi_{i+j} / d^2 = \mathbf{1}_d$ , and summing over  $A$  leaves  $Y$  and  $X$  uncorrelated. Hence, by identifying  $Z = A \times \tilde{Z}$  and  $q = (\mathbf{1}_d)_A \otimes \tilde{q}$ , the statement of the lemma follows.  $\square$

## V. $S$ IS THE ONLY CONTINUOUS ADDITIVE MONOTONE

Here we give a proof of Corollary 3 in the main text. This corollary follows immediately from the following lemma, which itself has Lemma 2 as its key ingredient.

**Lemma 9** (Properties of real and additive functions). *Let  $f$  be a real function on the set of all finite-dimensional density matrices which is continuous (on all subsets of density matrices of fixed dimension) and additive, i.e.  $f(\rho \otimes \sigma) = f(\rho) + f(\sigma)$ . Furthermore, suppose that  $f$  is a monotone with respect to transitions of the form (b) of Conjecture 1 in the main text, i.e. satisfaction of condition (b) implies that  $f(\rho) \leq f(\rho')$ . Then there exist a constant  $a \geq 0$  and dimension-dependent constants  $b_n \in \mathbb{R}$ , such that*

$$f(\rho) = a \cdot S(\rho) + b_n,$$

with  $n$  the Hilbert space dimension of  $\rho$ , and  $b_{m,n} = b_m + b_n$ .

*Proof.* For any density matrix  $\rho$  of dimension  $n$ , define the negentropy  $I(\rho) := \log n - S(\rho)$ . Let  $\rho, \rho'$  be full-rank density matrices of possibly different dimensions  $n, n'$  such that  $I(\rho) = I(\rho')$ , then

$$S(\rho \otimes \mathbf{1}_{n'}) = \log n - I(\rho) + \log n' = S(\rho' \otimes \mathbf{1}_n).$$

Let  $\epsilon > 0$ , and let  $\sigma_\epsilon$  be any full-rank state of size  $nn'$  such that  $\|\sigma_\epsilon - \rho \otimes \mathbf{1}_{n'}\| < \epsilon$  and  $S(\sigma_\epsilon) < S(\rho \otimes \mathbf{1}_{n'})$ , then  $S(\sigma_\epsilon) < S(\rho' \otimes \mathbf{1}_n)$ , hence Lemma 2 implies that there is some  $d \in \mathbb{N}$  such that  $\sigma_\epsilon \otimes \mathbf{1}_d \rightarrow \rho' \otimes \mathbf{1}_n \otimes \mathbf{1}_d$ , where “ $\rightarrow$ ” denotes that a transition of the form (b) is possible. Thus

$$f(\sigma_\epsilon \otimes \mathbf{1}_d) \leq f(\rho' \otimes \mathbf{1}_n \otimes \mathbf{1}_d),$$

and additivity of  $f$  yields  $f(\sigma_\epsilon) \leq f(\rho' \otimes \mathbf{1}_n)$ . Since  $\lim_{\epsilon \rightarrow 0} \sigma_\epsilon = \rho \otimes \mathbf{1}_{n'}$ , and since  $f$  is continuous, this implies that  $f(\rho \otimes \mathbf{1}_{n'}) \leq f(\rho' \otimes \mathbf{1}_n)$ . Reversing the roles of  $\rho$  and  $\rho'$  in the above argumentation gives the converse inequality, and hence  $f(\rho \otimes \mathbf{1}_{n'}) = f(\rho' \otimes \mathbf{1}_n)$ . Define the new real function  $j(\tau) := f(\mathbf{1}_n) - f(\tau)$ , where  $n$  is the dimension of the density matrix  $\tau$ , then  $j$  is also additive, and it vanishes on the maximally mixed states. Thus  $j(\rho) = j(\rho')$ .

In summary, we have shown that  $j$  is constant on the level sets of  $I$ . Thus, there is a real function  $g : [0, \infty) \rightarrow \mathbb{R}$  such that  $j(\rho) = g(I(\rho))$  for all  $\rho$ . Let  $x, y \in [0, \infty)$  with  $x < y$ , and let  $\rho_x, \rho_y$  be finite-dimensional full-rank density matrices of dimensions  $n_x, n_y$  with  $I(\rho_x) = x$  and  $I(\rho_y) = y$ . Then

$$\begin{aligned} g(x+y) &= g(I(\rho_x) + I(\rho_y)) = g(I(\rho_x \otimes \rho_y)) \\ &= j(\rho_x \otimes \rho_y) = j(\rho_x) + j(\rho_y) \\ &= g(I(\rho_x)) + g(I(\rho_y)) = g(x) + g(y). \end{aligned}$$

Furthermore,  $S(\rho_y \otimes \mathbf{1}_{n_x}) < S(\rho_x \otimes \mathbf{1}_{n_y})$ , hence there is some  $d \in \mathbb{N}$  such that  $\rho_y \otimes \mathbf{1}_{n_x} \otimes \mathbf{1}_d \rightarrow \rho_x \otimes \mathbf{1}_{n_y} \otimes \mathbf{1}_d$ , therefore  $j(\rho_y \otimes \mathbf{1}_{n_x} \otimes \mathbf{1}_d) \geq j(\rho_x \otimes \mathbf{1}_{n_y} \otimes \mathbf{1}_d)$ , and additivity implies  $j(\rho_y) \geq j(\rho_x)$ . It follows that  $g(y) \geq g(x)$ .

We thus see that  $g$  is both *additive* and *non-decreasing*, and

it is well-known (and easy to verify) that this implies that  $g(x) = ax$  for some  $a \geq 0$ , i.e.  $j(\rho) = aI(\rho)$ . Going back to the definition of  $f$ , this gives us

$$f(\rho) = aS(\rho) + b_n,$$

with  $n$  the dimension of  $\rho$  and  $b_n := f(\mathbf{1}_n) - a \log n$ . Finally, additivity of  $f$  and  $\mathbf{1}_{m,n} = \mathbf{1}_m \otimes \mathbf{1}_n$  imply  $b_{m,n} = b_m + b_n$ .  $\square$

Note that  $b_{m,n} = b_m + b_n$  does not automatically entail that  $b_m$  is proportional to  $\log m$  (and thus to  $S_0$ ): there are other well-known examples of functions on the integers which are additive in this sense.

- 
- [1] M. P. Müller, Physical Review X **8** (2018), 10.1103/physrevx.8.041051.  
 [2] A. Horn, Am. J. Math. **76**, 620 (1954).  
 [3] J. Schwinger, Proc. Natl. Ac. Sc. **46**, 570 (1960).  
 [4] R. F. Werner, J. Phys. A **34**, 7081 (2001).  
 [5] M. Klimesh, "Inequalities that collectively completely

- characterize the catalytic majorization relation," (2007), arXiv:0709.3680v1.  
 [6] S. Turgut, J. Phys. A **40**, 12185 (2007).  
 [7] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Yunger Halpern, Phys. Rep. **583**, 1 (2015).  
 [8] H. Wilming and R. Gallego, Phys. Rev. X **7**, 041033 (2017).





## 5.2 CATALYSIS FOR FLUCTUATION THEOREMS

In the last chapter, we have studied the question which state transitions are possible in the setting of correlated catalysis and we have seen that this setting is extremely powerful, with possible transitions essentially being constrained only by an increase of the von Neumann entropy (although we could prove this only up to the presence of a decohering environment). From a practical point of view, this is interesting primarily because states with similar entropy might have wildly different spectra, especially if the dimension of the system in question is large. For this reason, correlated catalysis allows one to modify the spectrum of a state significantly. It is natural, then, to look for tasks in which this ability can be put to use.

In this chapter we provide an very interesting example of the usefulness and counterintuitive power of correlated catalysis, by showing that the latter can be used to extract finite amounts of work per particle from a macroscopic system in thermodynamic equilibrium with a probability that is independent of the system size. This stands in stark contrast to thermodynamic intuition and bounds on work extraction that are implied by fluctuation theorems, in particular the Jarzynski equality.

## 5.2.1 Fluctuation theorems

The development of fluctuation theorems is relatively recent in the history of statistical mechanics [20, 21]. In their standard formulation, these theorems are concerned with establishing constraints on the possible fluctuations of thermodynamic quantities such as entropy production or work cost for systems that are driven out of equilibrium. In particular, in the setting of the seminal Crooks fluctuation theorem [146], one considers the probability of expending some amount of work in changing the Hamiltonian of a system over time, where this system is in contact with a heat bath and a) hence initially described by a thermal state and b) the trajectory of this system through its phase space – induced by an external driving that changes the Hamiltonian of the system from  $H_S^i$  initially to  $H_S^f$  at the end – is stochastic due to the system's interaction with the bath. Now, if we denote by  $P_+(w)$  the total probability of having implemented a trajectory with associated work *gain*  $w$  for a given external driving, and we denote by  $P_-(-w)$  the probability of having implemented a trajectory with associated work *cost*  $w$ , if we had started with the initial system Hamiltonian  $H_S^f$  and with the driving being reversed, then under very general conditions on the form of driving, Crook's fluctuation theorem states that

$$Z(\beta, H_S^i)P_+(w) = Z(\beta, H_S^f)e^{-\beta w}P_-(-w). \quad (5.2.1)$$

For example, if the driving was such that  $H_S^i = H_S^f$  (so that the process is cyclic), this would have meant that it is exponentially more likely to implement the process in that direction that incurs a work cost. One important implication of Crook's theorem is the so-called *Jarzynski equality*, which can be stated as [23]

$$\langle e^{\beta W} \rangle = e^{\beta \Delta F} \quad (5.2.2)$$

with  $W$  being the random variable that is the work extraction of the above process and  $\Delta F = F(\omega_\beta(H_S^f)) - F(\omega_\beta(H_S^i))$ . The Jarzynski equality is particularly noteworthy because in it, the expected work cost of a system that can drive a system arbitrarily far from equilibrium is constrained by the *equilibrium* properties of thermal states. As such, fluctuation theorems have much improved on researchers' access to the analytical study of far-from equilibrium processes.

### 5.2.2 *Bypassing fluctuation theorems with correlated catalysis*

In the publication presented below [3], we introduce correlated catalysis into the study of fluctuation theorems. Several works have generalized fluctuation theorems to quantum physics (see [24, 147, 148, 149, 150] and references therein). Here, the essentials of the process remain unchanged, with the difference that one commonly considers energy measurements at the beginning and end of an evolution, while the evolution itself is described as a unitary process that results from the change of the system (or system-bath) Hamiltonian due to external driving. We generalize this setting by allowing for the unitary channel between the two measurements to be replaced by a more general unitary that acts on both the system *and* an additional catalyst, under the usual requirement that the unitary leaves the average state of the catalyst unchanged.

Allowing for such catalytic evolutions clearly increases the possible trajectories on the system. The central result of the publication is to show that that not only does a Jarzynski equality-type constraint not hold anymore in this generalized setting, one can even bypass the constraints on the work distribution exhibited in Eq.(5.2.1)! Specifically, Eq. (5.2.2) implies that for macroscopic systems, the probability of extracting any finite amount of work per particle becomes exponentially small with the size of the system. In contrast, we provide an explicit construction for a process that uses correlated catalysis to achieve the extraction of a finite work per particle with a probability that can be brought, for any system size, arbitrarily close to 0.5, that is, to exponentially outperform the usual bound imposed by fluctuation theorems that do not admit catalysts. A good understanding of processes that achieve this work extraction could be of great operational use and also provide hints to the mechanisms underlying some classes of negentropic processes in nature.

The counterintuitive power of correlated catalysis resides both in the fact that the catalyst is itself a system that is far from equilibrium (one can show that catalysts in equilibrium could not be used to bypass Eq. (5.2.1)) and that the system is allowed to establish correlations with the catalyst. To better understand the sense in which establishing correlations provides a kind of resource, we also study scenarios in which a correlated catalyst is used to “engineer” global work distributions between many parties that locally only interact with the catalyst and find that strong correlations can be established, leading to interesting work distribution patterns in a many-player scenario. This concern with the power of correlations in the context of fluctuation theorems also reflects a growing theme in the quantum thermodynamics literature in which the role of correlations as a resource is examined theoretically [151, 152, 153, 154, 155, 156, 157] but also in experiments [31].

# By-passing fluctuation theorems with a catalyst

P. Boes<sup>1</sup>, R. Gallego<sup>1</sup>, N. H. Y. Ng<sup>1</sup>, J. Eisert<sup>1</sup>, and H. Wilming<sup>2</sup>

<sup>1</sup>Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany

<sup>2</sup>Institute for Theoretical Physics, ETH Zurich, 8093 Zurich, Switzerland

Fluctuation theorems impose constraints on possible work extraction probabilities in thermodynamical processes. These constraints are stronger than the usual second law, which is concerned only with average values. Here, we show that such constraints, expressed in the form of the Jarzynski equality, can be bypassed if one allows for the use of catalysts—additional degrees of freedom that may become correlated with the system from which work is extracted, but whose reduced state remains unchanged so that they can be re-used. This violation can be achieved both for small systems but also for macroscopic many-body systems, and leads to positive work extraction per particle with finite probability from macroscopic states in equilibrium. In addition to studying such violations for a single system, we also discuss the scenario in which many parties use the same catalyst to induce local transitions. We show that there exist catalytic processes that lead to highly correlated work distributions, expected to have implications for stochastic and quantum thermodynamics.

## 1 Introduction

Consider a physical system in thermal equilibrium with its environment. The second law of thermodynamics dictates that it is impossible to extract positive average work from this system using reversible processes that are cyclic in the Hamiltonian. More precisely, if the system's initial state is represented by a canonical ensemble and we consider many iterations of a probabilistic process during which the Hamiltonian of the system is varied but returned to the initial Hamiltonian at the end, then it holds that

$$\langle W \rangle \leq 0, \quad (1)$$

where  $\langle W \rangle$  is the average work extracted during the process. We will refer to (1) as the *Average Second Law (Av-SL)*.

However, there exist significantly stronger constraints on the possible extracted work in the above type of processes, namely those imposed by *fluctuation theorems* [1, 2, 3]. Indeed, using such theorems, one can show that the probability of extracting a finite amount of positive work per particle is exponentially suppressed with

the number of particles in a system [1]. Once these different types of constraints are recognized, an interesting question arises: What are physically meaningful settings in which the probabilistic constraints imposed by fluctuation theorems can be circumvented, while still respecting the Av-SL? In particular, do fluctuation theorems also hold when an additional, cyclically evolving auxiliary system is allowed for?

In this work, we present an answer to this question, by introducing a class of processes that generalize the above reversible processes, are physically well motivated, compatible with (1), and yet allow for the extraction of positive work per particle with a probability that is independent of system size. We do so via the notion of a *catalytic process*, in which we allow for the reversible process to not only act on the system as such, but additionally on an auxiliary system that can be initially prepared in an arbitrary state, but whose marginal state has to be left invariant by the process. Such catalysts are well-motivated – they allow a general description of thermodynamic processes in which the system may be interacting with some experimental apparatus (such as a quantum clock [4, 5]), however not extracting energetic/information resources from such an ancilla. In terms of our discussion of the Av-SL above, catalysts correspond to the cyclically evolving auxiliary system. Despite being studied frequently in resource-theoretic formulations of thermodynamics [6, 7, 8, 9], catalytic processes have never been studied in the context of fluctuation theorems until now. Furthermore, even in previous works of catalysis, the exact form of the catalyst is highly state-dependent and therefore rarely studied explicitly [6, 8]. In this work, we make progress in the significant gaps in the knowledge of catalysis, by presenting and discussing constructive examples of such catalytic processes in the framework where fluctuation theorems are commonly derived. We show that, by sharing the same catalyst, a group of agents can follow collective strategies to achieve highly correlated work-distributions. This makes these processes interesting for the field of *quantum and stochastic thermodynamics* and potentially also for certain negentropic processes in biology. On the overall, our work provides a rigorous footing for the further study of thermodynamical processes that systematically exploit the notion of *catalysis* in order to achieve certain patterns of work fluctuations in an environment that is governed by the Av-SL. Given the broad applicability of our results, we believe that the study of such processes will produce many further interesting results of both foundational and practical interest.

## 2 Setup

### 2.1 Formulation of the physical situation

We formulate our arguments and results in the language of quantum mechanics, but all of our results similarly apply to classical, stochastic systems. We consider the setting depicted in Fig. 1: A  $d$ -dimensional system  $S$  with Hamiltonian  $H = \sum_{i=1}^d E_i |E_i\rangle\langle E_i|$  is initialized in the Gibbs state

$$\omega_\beta(H) := \frac{e^{-\beta H}}{Z(\beta, H)},$$

where  $Z(\beta, H) := \text{Tr}(e^{-\beta H})$ . This state describes a system initially in thermal equilibrium with its environment at inverse temperature  $\beta := 1/(k_B T)$ . An agent (some experimenter) first performs an energy measurement on this system which produces a measurement outcome  $E_i$ . According to quantum mechanics, the post-measurement state is described by the density matrix  $|E_i\rangle\langle E_i|$ . The agent then performs a physical operation on the system which does not depend on the outcome of the measurement. Such an operation can always be represented by a general quantum channel  $\mathcal{C}$  (i.e., a trace-preserving, completely positive map that takes density matrices to density matrices) applied to the post-measurement state. This operation is then followed by a second energy measurement with respect to the same Hamiltonian with outcome  $E_f$ <sup>1</sup>. This procedure results in a channel-dependent joint distribution  $P(E_f, E_i) = P(E_f|E_i)P(E_i)$ . In general, a given quantum channel may be realized in different ways. Whether the change of energy  $E_f - E_i$  can be interpreted as work from a thermodynamic point of view will depend on how exactly the quantum channel  $\mathcal{C}$  was physically realized. We will assume that this is the case in the following, but will comment on this assumption again later on. In particular, we can then define the work distribution  $P$  for the above process as

$$P(W) := \sum_{i,f} P(E_f, E_i) \delta(W - (E_i - E_f)),$$

where  $\delta$  is the Dirac delta distribution. We are interested in investigating possible distributions  $P(W)$  that arise from different channels  $\mathcal{C}$ . To do so, it is useful to note the relation

$$\langle e^{\beta W} \rangle = \sum_j \frac{e^{-\beta E_j}}{Z_H} \langle E_j | \mathcal{C}[\mathbf{I}] | E_j \rangle, \quad (2)$$

which is straightforwardly derived using the above definitions, where  $\mathbf{I}$  denotes the identity matrix.

In the standard setting of *Tasaki-type fluctuation theorems*,  $\mathcal{C}$  is considered to be a unitary channel  $\mathcal{C}[\cdot] = U(\cdot)U^\dagger$ , since these are generated by changing the Hamiltonian over time [3]. For such channels, (2) becomes

$$\langle e^{\beta W} \rangle = 1, \quad (3)$$

<sup>1</sup>It is possible to extend the setup and our further results to the more general case of different Hamiltonians for the initial and final measurement. We present our results within this restricted settings for conceptual and notational simplicity.

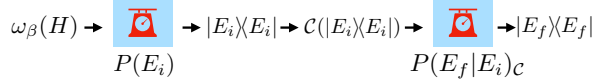


Figure 1: The basic setup for all processes in this work: An agent with access to a system  $S$  equipped with Hamiltonian  $H$  that is assumed to be initially in thermal equilibrium with a heat bath at inverse temperature  $\beta$  samples from  $S$  (by measuring in the energy basis), then implements a process that maps the post-measurement state  $|E_i\rangle\langle E_i|$  to  $\mathcal{C}(|E_i\rangle\langle E_i|)$ , where  $\mathcal{C}$  is a quantum channel. Finally, the agent repeats the energy measurement on  $S$  with respect to the same Hamiltonian  $H$ .

which is the well-known Jarzynski equality (JE) for cyclic, reversible processes [1]. Eq. (3) is strictly stronger than (1), the latter being implied by (3) via Jensen's inequality.

### 2.2 No macroscopic work

One of the reasons for the importance of the JE derives from the fact that it gives strong bounds on the possibility of extracting work from a large system in a thermal state [10, 11, 12]. To see this, let  $S$  be an  $N$ -particle system and define the probability of extracting work  $w$  per particle as

$$p(w) := P(wN).$$

Plugging this into (3) yields that for any  $\epsilon > 0$ ,

$$1 = \langle e^{\beta W} \rangle = \sum_w e^{\beta w N} P(wN) \geq e^{\beta \epsilon N} \sum_{w \geq \epsilon} p(w),$$

which implies that events which extract significant positive work per particle from a macroscopic system at equilibrium are exponentially unlikely in  $N$ . For later use, we formalize this property.

**Definition 1** (No macroscopic work). *Given a sequence of  $N$ -particle systems initially at thermal equilibrium with inverse temperature  $\beta$  and channels  $\mathcal{C}$  (implicitly depending on  $N$ ), we say that the processes represented by  $\mathcal{C}$  fulfill the no macroscopic work (NMW) condition if the probability of an event extracting work per particle larger or equal than  $\epsilon$  is arbitrarily small as  $N \rightarrow \infty$ ,*

$$\lim_{N \rightarrow \infty} p(w \geq \epsilon) := \lim_{N \rightarrow \infty} \sum_{w > \epsilon} p(w) = 0.$$

As is clear from the above, channels that satisfy the JE, such as unitary channels, also satisfy NMW and Av-SL. We now turn to investigate violations of JE and NMW for non-unitary channels.

## 3 Violations of NMW and JE

The first main result of this work is to introduce a physically motivated family of channels  $\mathcal{C}$  that violates both

NMW and JE, but respects the Av-SL. To aid comparison, we first briefly discuss other generalizations of the standard setting to non-unitary channels (see also Refs. [13, 14]).

### 3.1 Violating JE with non-unitary channels

It is easy to see from (2) that a more general class of channels that satisfy the JE are *unital* channels, that is, channels that satisfy  $\mathcal{C}[\mathbf{I}] = \mathbf{I}$ . Consequently, neither JE, nor in turn NMW or Av-SL can be violated in settings which give rise to a unital channel. However, once this condition on unitality is relaxed, it becomes easy to violate JE on a formal level. For example, consider the fully-thermalizing channel that maps every input state to the thermal state  $\omega_\beta(H)$ , in other words  $\mathcal{C}(\cdot) = \omega_\beta(H)$ . This channel always violates the JE whenever  $\omega_\beta(H) \neq \mathbf{I}/d$ . It is, however, not clear how the energy-fluctuations can be interpreted as *work* in this example, since thermalizing processes usually occur due to contact with a heat bath, in which case one would naturally interpret the changes of energy on the system being due to heat. Thus, while it is trivial to formally violate JE, it is not obvious whether it is possible to do so in a physically meaningful and operationally useful manner. Nevertheless, in Appendix A, we show that the fully-thermalizing channel, in fact any channel with the thermal state as a fixed point, cannot violate the NMW condition for typical many-body systems, even if they may violate (3). This means that, even if one interprets energy fluctuations as work, one still could not use the thermalizing channel to extract macroscopic amounts of work from a many-body system.

### 3.2 Violations of NMW and JE via $\beta$ -catalytic channels

The above findings raise the important question whether there exist channels for which the above procedure leads to a violation of NMW (and hence JE), while still respecting the Av-SL and allowing for the interpretation of the random variable  $W$  as work extracted from  $S$ . Such channels, if they exist, promise to be of great interest because they could allow for a systematic exploitation of relatively likely events extracting work from heat baths. The first result of this work is to answer this question affirmatively. To this end, we define the notion of a  $\beta$ -catalytic channel.

**Definition 2** ( $\beta$ -catalytic channel). *A completely positive, trace-preserving map  $\mathcal{C}$  is a  $\beta$ -catalytic channel on  $S$ , if there exists a quantum state  $\sigma_C$  on a system  $C$  with Hamiltonian  $H_C$ , together with a unitary  $U$  such that  $[\sigma_C, H_C] = 0$  and*

$$\begin{aligned} \mathcal{C}(\cdot) &= \text{Tr}_C(U(\cdot \otimes \sigma_C)U^\dagger), \\ \text{s.t. } \text{Tr}_S(U(\omega_\beta(H) \otimes \sigma_C)U^\dagger) &= \sigma_C. \end{aligned} \quad (4)$$

Before stating our first main result, let us make some comments about this definition. First of all, we already assumed that the initial and final Hamiltonian coincides.

This means that while during the process,  $C$  may couple system and catalyst for example by introducing interaction terms  $H_{SC}$ , nevertheless at the end of the process, the channel must also turn off such interaction terms. Secondly, note that  $\beta$ -catalytic channels describe reversible processes, in the sense that they do not change the entropy of the joint-system  $SC$  and can be undone by acting on this joint-system by a unitary process. We refer to the system  $C$  as being the ‘‘catalyst’’, understanding that it may be some by-stander system involving additional degrees of freedom. This terminology is motivated by the fact that, on average, i.e., if we do not condition on the outcomes of the energy measurements, then  $C$  is returned, at the end of the procedure, to its original state. It can therefore be reused for further rounds of the protocol with *new* copies of  $S$ . Note, however, that the invariance of the reduced state on  $C$  under the channel is required not for all initial states of  $S$ , but *only* for  $\omega_\beta(H)$ . As such,  $\beta$ -catalytic channels depend on  $\beta$  and  $H$  through the second condition.

While Definition 2 does not require the catalyst to be uncorrelated with  $S$  at the end of the protocol, and in this sense goes beyond the conventional notion of catalysis discussed in the resource-theoretic literature on quantum thermodynamics [6, 7], the more general notion of catalysis that we employ here is receiving increasing interest in quantum thermodynamics, where it was shown to single out the quantum relative entropy, free energy and von Neumann entropy [15, 8, 16], to be useful in the context of algorithmic cooling [16, 17] and to show the energetic instability of passive states [18]. Finally, let us briefly comment on the interpretation of the random variable  $W$  as work in the setting of  $\beta$ -catalytic channels and the role of the Hamiltonian of the catalyst. Since the process on  $C$  and  $S$  is unitary, it is meaningful to denote the total changes of energies of the two systems as work measured by a two-point measurement scheme on each system. This gives rise to a joint-distribution of work on the two systems  $P(W^{(S)}, W^{(C)})$ . The probability distribution of work  $P(W)$  discussed above then simply corresponds to the marginal distribution  $P(W^{(S)})$  on  $S$ . Importantly, this distribution is independent of the Hamiltonian on  $C$  (see Sec. G in the Appendix). In particular, we can assume that the catalyst has trivial Hamiltonian  $H_C = 0$ , which in turn implies  $[\sigma_C, H_C] = 0$  for any  $\sigma_C$ . It is then clear that no energy flows from the catalyst to the system, not even probabilistically. For the rest of the article, we hence assume that  $H_C = 0$ .

Given these constraints, it may, at first glance, be unclear how such a catalyst would offer any advantage to violating JE. For instance, one apparent way to make use of the catalyst is to perform a controlled unitary on  $S$ , conditioned on  $C$ : For some  $\sigma_C = \sum_i p_i |i\rangle\langle i|$ , one uses a unitary in Eq. (4) of the form

$$U_{SC} := \sum_i U_i \otimes |i\rangle\langle i|_C.$$

This special case of  $\beta$ -catalytic channels by construction produces random unitary channels [19, 16] on  $S$ , which



have the form  $\mathcal{C}_{\text{RU}}(\cdot) = \sum_i p_i U_i(\cdot) U_i^\dagger$ . But random unitary channels are always unital, and therefore automatically satisfy JE.

In the following, we show that there exist non-unital  $\beta$ -catalytic channels that allow for a meaningful violation of both NMW and JE, while at the same time they always respect the Av-SL. To see the latter, we note that these channels necessarily increase the von Neumann entropy of the input Gibbs state. This follows from the sub-additivity of entropy and the fact that  $C$  remains locally unchanged. Now, since  $\omega_\beta(H)$  is the state with the least energy given a fixed entropy [20, 21], then we also have that

$$\text{Tr}(HC(\omega_\beta(H))) \geq \text{Tr}(H\omega_\beta(H))$$

which is just the Av-SL, concomitant with the findings of Ref. [9]. We stress that despite this property,  $\beta$ -catalytic channels are in general *not* unital. It remains to be shown that  $\beta$ -catalytic channels that violate JE and NMW do exist. We first show that JE can be violated already with small quantum systems, and then turn to the violation of NMW for macroscopic many-body systems with physically realistic Hamiltonians.

*Microscopic violation of JE.* As a toy-like example of violating the JE with  $\beta$ -catalytic channels, we consider a system with three states – two degenerate (but distinguishable) ground states and an excited state with energy  $E$ . As catalyst, we consider a system with two states and the unitary is a simple permutation between two pairs of energy eigenvalues of the joint system (for details, see App. B). It is straightforward to compute the probability distribution of work for such small systems, which in this case leads to

$$\langle e^{\beta W} \rangle = \frac{Z + 5 + 2(Z - 2)(Z - 1)}{Z(Z + 1)} \geq 1,$$

where  $Z = 2 + e^{-\beta E}$  is the partition function of the system and we used  $2 \leq Z \leq 3$ . We hence find  $\langle e^{\beta W} \rangle > 1$  whenever  $E > 0$  (since then  $Z < 3$ ) and we obtain a moderate maximum violation in the limit  $E \rightarrow \infty$  given by  $\langle e^{\beta W} \rangle = 7/6$ .

*Macroscopic violation of NMW condition.* We now show that one can violate the NMW principle using catalysts.

**Proposition 1** (Violation of no macroscopic work with catalysts). *Let  $(S^{(N)})_N$  be a sequence of  $N$ -particle locally interacting lattice systems with Hamiltonian  $H^{(N)}$  that satisfy mild assumptions. Then, for sufficiently large  $N$ , there exist values of  $\epsilon > 0$ , such that*

$$p(w \geq \epsilon) \quad (5)$$

*can be brought arbitrarily close to  $\frac{1}{2}$  with  $\beta$ -catalytic channels.*

We provide a proof and full statement of the assumptions in Appendix D. Our assumptions are satisfied by

typical many-body Hamiltonians with energy windows in which the density of states grows exponentially [22].

While the formal proof of Proposition 1 is given in the Appendix, the idea behind it is simple and we sketch it here on a higher level. For a given  $N$ , let  $e^{(N)}$  denote the mean energy per particle of an  $N$ -particle system that satisfies our assumptions. In the proof, we show that for systems that satisfy the above assumptions and any  $\delta > 0$ , there exists an  $N$  and a  $\beta$ -catalytic channel  $\mathcal{C}$  such that

$$\mathcal{C}(\omega_\beta) \approx_\delta \frac{1}{2} |E_- \rangle \langle E_-| + \frac{1}{2} \tau, \quad (6)$$

where  $\approx_\delta$  denotes equality of the states on LHS and RHS up to  $\delta$  in trace distance,  $|E_- \rangle$  is some eigenvector of  $H$  with  $E_- < e^{(N)}N$  and  $\tau$  is some other “fail”-state the details of which are irrelevant. We can interpret Eq. (6) as describing the approximation of a work extraction protocol that results in the state  $|E_- \rangle$  with probability  $1/2$ . Now, as the result of standard concentration bounds, for large  $N$  the mass of the thermal state  $\omega_\beta$  will be highly concentrated around energy  $e^{(N)}N$ . This implies that every time the above work extraction protocol succeeds to prepare the ground state, for sufficiently high values of  $N$  the extracted work per particle is arbitrarily close to  $\epsilon \equiv e^{(N)} - E_-/N$ , leading to the statement of Prop. 1.

We note that it is remarkable that catalytic channels, which are guaranteed to satisfy the Av-2nd law, allow for the preparation of states like the one described in Eq. (6), in which a pure low-energy state carries much of the weight, from a thermal state. Indeed, it has recently been conjectured that with the help of catalysts *any* state transition between full-rank states that increases the entropy is possible [9], a statement known as the *catalytic entropy conjecture*. Prop. 1, and in particular the ability to prepare the state in Eq. (6), further supports this conjecture, which has not been proven so far (even though strong evidence has been established).

Similar results as above also apply to the case in which the initial state of the system is described by a *micro-canonical ensemble* rather than the Gibbs state, highlighting a similar contrast to fluctuation theorem results in the micro-canonical regime [23]. For detailed discussions and proves of corresponding statements in this regime, see Appendix C.

One may wonder whether the creation of correlations between system and catalyst is in fact necessary to violate the NMW principle. This is indeed true, when one simply forces the catalyst to remain uncorrelated in the definition of  $\beta$ -catalytic channels. A proof of this statement along with further discussion on this problem can be found in Appendix I. Interestingly, such processes at the same time allow for a violation of the Jarzynski equality. A particular example is given by the fully thermalizing channel, which can be realized using a catalyst that is simply a copy of the Gibbs state of the system and the unitary simply swapping the system and catalyst.

*Required size of the catalyst.* Proposition 1 not only shows that there exist catalytic procedures that allow an agent to bypass the work extraction bounds imposed by the

JE – the violation of JE is in fact exponential in the system size. In particular, (5) implies that there exist values  $\epsilon > 0$ , such that

$$\langle e^{\beta W} \rangle \geq \frac{1}{2} e^{\beta N \epsilon} \gg 1$$

in the limit of large  $N$ . It is natural to wonder how far the JE can be violated and how big the catalyst has to be to realize a certain violation. This is clarified by the following result.

**Proposition 2** (Bound on violation of JE). *Let  $\mathcal{C}$  be any  $\beta$ -catalytic channel with  $d_C = \dim(H_C)$ . Then,*

$$\begin{aligned} \langle e^{\beta W} \rangle &\leq \min\{d_C \|\sigma\|_\infty, d \|\omega_\beta(H)\|_\infty\} \\ &\leq \min\{d_C, d\}, \end{aligned}$$

where  $\|\cdot\|_\infty$  denotes the  $\infty$ -norm, which, for density matrices, equals the largest absolute value of the input's eigenvalues.

This proposition, the simple proof of which is given in Appendix F, shows that in order to extract a growing amount of work from a single run of a process, an external agent will have to be able to prepare a state  $\sigma$  on a growing auxiliary system and, more importantly, also have control over the increasingly large joint system. Hence, in practice, the ability to violate JE will still be constrained by operational limitations. To illustrate the implications of Prop. 2, let us show how it immediately implies a bound on  $P(W)$ . As noticed when deriving the NMW principle, for any  $\epsilon \geq 0$  we have

$$\langle e^{\beta W} \rangle \geq P(W \geq \epsilon) e^{\beta \epsilon}.$$

Hence, Prop. 2 implies

$$P(W \geq \epsilon) \leq d_C \|\sigma\|_\infty e^{-\beta \epsilon}.$$

In particular this means that to extract a macroscopic amount of work,  $W \geq wN$ , with finite probability,  $d_C$  has to grow exponentially with  $N$  (note that  $\|\sigma\|_\infty \leq 1$ ).

## 4 Multi-partite work extraction

As emphasized before, even though the state of the catalyst remains unchanged in a catalytic process, in general it builds up correlations with the system. We now show that the correlations established between catalyst and system allow for processes in which many agents re-use the same catalyst to obtain highly inter-correlated work distributions.

Consider  $n$  agents, each with identical systems  $S_i, i \in \{1, \dots, n\}$  that are initialized in the Gibbs state  $\omega(\beta, H)$ . For a given  $\beta$ -catalytic channel  $\mathcal{C}$  with state  $\sigma$  on the catalyst, consider the following protocol: Agent 1 runs the standard process from Fig. 1 using the catalyst and hence implementing  $\mathcal{C}$  between the two measurements. After the procedure, she then passes  $C$  on to agent 2 who repeats

this process, and so on, until the last agent has received  $C$  and performed the process. From the catalytic nature of  $\mathcal{C}$ , it is clear that, for each agent, the same marginal distribution of work is obtained. However, the joint work distribution for all agents will be correlated, due to individual correlations between each  $S_i$  with  $C$ . We now show that the agents can use these correlations to systematically achieve certain global work distributions. Using the same notation as before, let  $p(w_1, \dots, w_n)$  denote the global distribution over the extracted work per particle, assuming that all  $S_i$  are copies of the same  $N$ -particle system. We have the following, proven in Appendix E.

**Proposition 3** (Multiple agents). *Let each  $\{S_i\}_{i=1}^n$  be a sequence of  $N$ -particle systems that satisfy the conditions of Proposition 1. Then, for sufficiently large  $N$ , there exists an  $\epsilon > 0$ , such that*

$$\begin{aligned} p(\epsilon, -\epsilon, \epsilon, -\epsilon, \dots) &= \lambda, \\ p(-\epsilon, \epsilon, -\epsilon, \epsilon, \dots) &= 1 - \lambda, \end{aligned} \quad (7)$$

where  $\lambda$  can be brought arbitrarily close to  $1/2$  using a sequence of  $\beta$ -catalytic channels on  $S_i$  and  $C$ .

While (7) is clearly consistent with (1), this proposition shows that the agents can achieve joint work distributions that are strongly correlated and in which subsets of agents, in the above proposition one half of them, can violate JE arbitrarily, at the cost of the other half. Such distributions of work could, for example, be of interest in situations where the target is to maximize the probability that a subset of players extracts a positive amount work, at the ready cost of the others, for instance in order to surpass an activation energy. Importantly, the size of the catalyst needed to realize the distribution (7) is fixed, i.e., it does not scale with the number of agents  $n$ .

Proposition 3 shows the existence of catalytic processes that produce very interesting global work distributions. This naturally raises the question what other global distributions can be obtained in a setting without making the size of the catalyst depend on the number of rounds. Our results, however, already imply that not every distribution compatible with the Second Law can be obtained in such a way. For instance, Proposition 2 implies that the distribution

$$p(\epsilon, \epsilon, \epsilon, \epsilon, \dots) = p(-\epsilon, -\epsilon, -\epsilon, -\epsilon, \dots) \approx 1/2$$

cannot be obtained via  $\beta$ -catalytic channels, since otherwise there would exist a catalyst of fixed size that would allow, for any  $n$ , the total work  $W = n\epsilon$  to be extracted with probability approximately  $1/2$ , in violation of Proposition 2.

## 5 Summary and future work.

In this work we have studied work extraction protocols from states at thermal equilibrium. We significantly expand the common setting of fluctuation theorems under

cyclic, reversible processes by introducing a catalyst—an additional system which, on average, remains unchanged after the protocol and can thus be re-used. This extension enables for distributions of work extraction that are not attainable without a catalyst. More precisely, one can bypass the stringent conditions imposed by the JE, achieving positive work per particle with high probability, even for macroscopic systems. Furthermore, it allows for interesting, correlated work distributions when many agents use the same catalyst.

Our constructions illustrate in a striking way that the absence of correlations, sometimes referred to as ‘stochastic independence’, can also be a powerful thermodynamic resource [24]. This complements findings where the initial presence of correlations between a system and an ancilla are used to bypass the standard constraints imposed by fluctuation theorems [25, 26]. We discuss the connection of our work to these findings in more detail in Appendix H. We believe that the further study of work distributions that can be obtained by collaborating agents by means of  $\beta$ -catalytic channels will yield both foundational and practical insights.

We further believe that it is an interesting open problem to study how the size of the catalyst has to scale if one wishes to maximize the probability to extract a certain amount of work. For example, in the context of a many-body system one might be content with extracting only an amount of work of the order of  $\sqrt{N}$  if in exchange for that one can either increase the probability for it to happen significantly or can reduce the size of the catalyst considerably (and hence the complexity of the unitary required to be implemented).

It would be interesting to understand the relation between our results and a more generalized type of JE in the presence of information exchange [27], for example in a Maxwell demon scenario. In particular, in Ref. [28] it was also demonstrated that by using feedback control, one may also violate JE while respecting the Av-SL. More generally, our results also raise the question whether other phenomena—usually described as forbidden by the second law, or as occurring with vanishing probability—can be made to occur with high probability using catalysts. For example, is it possible to reverse the mixing process of two gases or induce heat flow from a cold to a hot system with finite probability in macroscopic systems? The techniques developed in this work provide a promising ansatz for the study of this and similar questions.

*Acknowledgements.* We thank Markus P. Müller and Alvaro M. Alhambra for valuable discussions and anonymous referees for interesting comments. P. B. acknowledges support from the John Templeton Foundation. H. W. acknowledges support from the Swiss National Science Foundation through SNSF project No. 200020\_165843 and through the National Centre of Competence in Research *Quantum Science and Technology* (QSIT). N. H. Y. N. acknowledges support from the Alexander von Humboldt Foundation. R. G. has been supported by the DFG (GA 2184/2-1). J. E. acknowledges

support by the DFG (FOR 2724), dedicated to quantum thermodynamics, and the FQXi.

## References

- [1] C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).
- [2] G. E. Crooks, *J. Stat. Phys.* **90**, 1481 (1998).
- [3] H. Tasaki, ArXiv e-prints (2000), arXiv:1303.6393 .
- [4] P. Erker, M. T. Mitchison, R. Silva, M. P. Woods, N. Brunner, and M. Huber, *Phys. Rev. X* **7**, 031022 (2017).
- [5] M. P. Woods, R. Silva, and J. Oppenheim, *Ann. Hen. Poin.* **20**, 125 (2019).
- [6] F. G. S. L. Brandão, M. Horodecki, N. H. Y. Ng, J. Oppenheim, and S. Wehner, *PNAS* **112**, 3275 (2015).
- [7] N. H. Y. Ng, L. Mančinska, C. Cirstoiu, J. Eisert, and S. Wehner, *New J. Phys.* **17**, 085004 (2015).
- [8] M. P. Müller, *Phys. Rev. X* **8** (2018), 10.1103/physrevx.8.041051.
- [9] P. Boes, J. Eisert, R. Gallego, M. P. Mueller, and H. Wilming, *Phys. Rev. Lett.* **122**, 210402 (2019).
- [10] C. Jarzynski, *Annu. Rev. Condens. Matter Phys.* **2**, 329 (2011).
- [11] V. Cavina, A. Mari, and V. Giovannetti, *Scientific Rep.* **6**, 29282 (2016).
- [12] O. Maillet *et al.*, *Phys. Rev. Lett.* **122**, 150604 (2019).
- [13] A. E. Rastegin, *J. Stat. Mech.* **2013**, P06016 (2013).
- [14] A. E. Rastegin and K. Życzkowski, *Phys. Rev. E* **89**, 012127 (2014).
- [15] H. Wilming, R. Gallego, and J. Eisert, *Entropy* **19**, 241 (2017).
- [16] P. Boes, H. Wilming, R. Gallego, and J. Eisert, *Phys. Rev. X* **8**, 041016 (2018).
- [17] Á. M. Alhambra, M. Lostaglio, and C. Perry, *Quantum* **3**, 188 (2019).
- [18] C. Sparaciari, D. Jennings, and J. Oppenheim, *Nat. Commun.* **8**, 1895 (2017).
- [19] K. M. R. Audenaert and S. Scheel, *New J. Phys.* **10**, 023011 (2008).
- [20] A. Lenard, *J. Stat. Phys.* **19**, 575 (1978).
- [21] W. Pusz and S. L. Woronowicz, *Comm. Math. Phys.* **58**, 273 (1978).
- [22] K. Huang, *Statistical mechanics* (Wiley, 1987).
- [23] P. Talkner, P. Hänggi, and M. Morillo, *Phys. Rev. E* **77**, 051131 (2008).
- [24] M. Lostaglio, M. P. Müller, and M. Pastena, *Phys. Rev. Lett.* **115**, 150402 (2015).
- [25] T. Sagawa and M. Ueda, *Phys. Rev. Lett.* **109**, 180602 (2012).
- [26] T. Sagawa and M. Ueda, *New J. Phys.* **15**, 125012 (2013).
- [27] T. Sagawa and M. Ueda, *Phys. Rev. Lett.* **109**, 180602 (2012).
- [28] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano, *Nature Phys.* **6**, 988 (2010).



- [29] M. Horodecki and J. Oppenheim, *Nature Comm.* **4**, 2059 (2013).
- [30] C. Perry, P. Ćwikliński, J. Anders, M. Horodecki, and J. Oppenheim, *Phys. Rev. X* **8**, 041049 (2018).
- [31] P. Faist, J. Oppenheim, and R. Renner, *New J. Phys.* **17**, 043003 (2015).
- [32] C. Gogolin, M. P. Müller, and J. Eisert, *Phys. Rev. Lett.* **106**, 040401 (2011).
- [33] A. Anshu, *New J. Phys.* **18**, 083011 (2016).
- [34] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).
- [35] S. Goldstein, T. Hara, and H. Tasaki, ArXiv e-prints (2013), [arXiv:1303.6393](https://arxiv.org/abs/1303.6393).
- [36] M. Kliesch, C. Gogolin, M. J. Kastoryano, A. Riera, and J. Eisert, *Phys. Rev. X* **4**, 031019 (2014).
- [37] J. Watrous, *The theory of quantum information* (Cambridge University Press, 2018).
- [38] M. Perarnau-Llobet, E. Bäumer, K. V. Hovhannisyan, M. Huber, and A. Acin, *Phys. Rev. Lett.* **118**, 070601 (2017).
- [39] J. V. Koski, V. F. Maisi, T. Sagawa, and J. P. Pekola, *Phys. Rev. Lett.* **113**, 030601 (2014).

## A NMW for Gibbs preserving maps

Thermalizing quantum maps, in particular those studied in the resource theoretic framework, are maps that model the evolution of a non-equilibrium quantum state as it exchanges heat with its surrounding thermal bath. Several variants of these maps exist [29, 6, 7, 30, 31], but a common feature is that they are *Gibbs preserving (GP)*, namely that the Gibbs canonical state is a fixed point of such maps. Thermalizing maps are often viewed as “free operations” in a resource theoretic context, since they allow only for heat (instead of work) exchange with an environment in thermal equilibrium. In this section, we demonstrate two things: First, that even such thermodynamically “cheap” channels may violate the JE very strongly, due to non-unitality. Secondly, that they cannot be used to violate the NMW condition. A diagrammatic overview over the various properties of channels with respect to JE and NMW is given in Fig. 2.

We now turn to the first point. Given a  $d$ -dimensional system  $S$  with Hamiltonian  $H$ , the violation of JE can be calculated for the thermalizing channel as

$$\begin{aligned} \langle e^{-\beta W} \rangle &= d \sum_j \frac{e^{-\beta E_j}}{Z_H} \langle E_j | \mathcal{C}[\mathbf{I}/d] | E_j \rangle, \\ &= d \sum_j \frac{e^{-\beta E_j}}{Z_H} \langle E_j | \omega_\beta(H) | E_j \rangle = \frac{d}{d_{\text{eff}}}, \end{aligned}$$

where  $d_{\text{eff}} := 1/\text{Tr}(\omega_\beta(H)^2)$  is known as the effective dimension [32] of the thermal state. One sees from the above that JE is always violated for  $\beta > 0$ , since  $d_{\text{eff}} \leq d$ , with equality only when  $\omega_\beta(H) = \mathbf{I}/d$  is maximally mixed. For  $N$  non-interacting i.i.d. systems, both  $d$  and  $\text{Tr}(\rho^2)$  scale exponentially with  $N$ , leading to an exponential violation in  $N$  for JE.

Turning to the second point, one may wonder how this notion of thermodynamically free channels can be reconciled with the fact that JE is violated. However, note that in the standard JE setting, the work variable is traditionally defined in terms of a fluctuating (measured) energy difference in the system, and does not inherently distinguish between work and heat contributions – unlike resource-theoretic settings where heat flow is allowed for free, but measurements incur a thermodynamic cost. Here, we consider an operationally more meaningful characterization (NMW as defined in Def. 1 of the main text), and show that NMW cannot be violated using channels that preserve the Gibbs state in generic many-body systems. The only assumptions that we make are that i) the system has uniformly bounded, local interactions on a  $D$ -dimensional regular lattice and ii) a finite correlation length, i.e., the temperature is non-critical.

**Lemma 3** (Non-violation of NMW for Gibbs-preserving maps). *No channel  $\mathcal{E}$  that preserves the Gibbs state can violate NMW for locally interacting many-body systems at a non-critical temperature.*

*Proof.* We aim at showing that for any  $a > 0$ ,  $p(w \geq a) = p(W \geq aN) \rightarrow 0$  as  $N \rightarrow \infty$ . The

basic idea behind our proof is to make use of typicality. Let  $e^{(N)}$  denote the energy density of the  $N$ -particle system and denote by  $\Pi_\delta^{(N)}$  the projector onto energy eigenstates with energies in the interval  $T_{N,\delta} := [(e^{(N)} - \delta)N, (e^{(N)} + \delta)N]$ . Finally, denote by  $p(\cdot)$  the initial probability distribution of energy of the thermal state  $\tau_S^{(N)}$ , e.g., the probability that the initial energy measurement yields  $E_i \in T_{N,\delta}$  is given by

$$p(T_{N,\delta}) := \text{Tr} \left( \tau_S^{(N)} \Pi_\delta^{(N)} \right).$$

A theorem by Anshu [33] shows that under the given conditions most weight of the thermal state  $\tau_S^{(N)}$  of the  $N$ -particle system is contained in a typical subspace. More precisely, for a many-body system described by a  $D$ -dimensional lattice, there exist constants  $C, K > 0$  such that for any  $\delta > 0$  we have

$$p(T_{N,\delta}) \geq 1 - Ce^{-\frac{(\delta^2 N)^{\frac{1}{1+D}}}{K}}. \quad (8)$$

This is equivalent to saying that

$$p(T_{N,\delta}^c) \leq Ce^{-\frac{(\delta^2 N)^{\frac{1}{1+D}}}{K}},$$

where  $T_{N,\delta}^c = \mathbb{R} \setminus T_{N,\delta}$ . In particular, in the case of  $D = 0$ , i.e.,  $N$  non-interacting systems, we find the usual scaling obtained from Hoeffding's inequality. In the following, for simplicity of notation, we write  $\sigma_1 = \tau_S^{(N)}$  and consider the normalized state  $\sigma_2$  obtained by restricting  $\tau_S^{(N)}$  to the subspace  $\Pi_\delta^{(N)}$  as

$$\sigma_2 := \frac{\Pi_\delta^{(N)} \tau_S^{(N)}}{p(T_{N,\delta})}.$$

Let us further write  $\mathcal{E}(\sigma_{1(2)}) = \sigma'_{1(2)}$ , where  $\sigma'_1 = \sigma_1$  by assumption. Since the trace distance  $d(\rho_1, \rho_2) := \frac{1}{2} \text{Tr}(|\rho_1 - \rho_2|)$  fulfills the data processing inequality,

$$d(\sigma_1, \sigma'_2) = d(\sigma'_1, \sigma'_2) \leq d(\sigma_1, \sigma_2) = p(T_{N,\delta}^c).$$

Using the operational meaning of trace distance  $d(\rho_1, \rho_2) = \max_{0 \leq M \leq I} |\text{Tr}(M(\rho_1 - \rho_2))|$  [34], this means that

$$|\text{Tr}(\Pi_\delta^{(N)} \sigma_1) - \text{Tr}(\Pi_\delta^{(N)} \sigma'_2)| \leq p(T_{N,\delta}^c) \quad (9)$$

and, in turn,

$$\text{Tr}(\Pi_\delta^{(N)} \sigma'_2) \geq p(T_{N,\delta}) - p(T_{N,\delta}^c) = 1 - 2p(T_{N,\delta}^c) \quad (10)$$

To see this, note that (10) follows from (9) directly if  $\text{Tr}(\Pi_\delta^{(N)} \sigma'_2) \leq \text{Tr}(\Pi_\delta^{(N)} \sigma_1)$ , and as

$$\text{Tr}(\Pi_\delta^{(N)} \sigma'_2) > \text{Tr}(\Pi_\delta^{(N)} \sigma_1) \geq \text{Tr}(\Pi_\delta^{(N)} \sigma_1) - p(T_{N,\delta}^c)$$

otherwise. This means that, conditioned on the fact that the initial state was within the typical energy

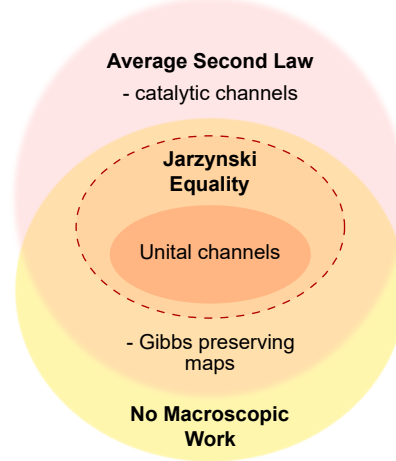


Figure 2: A summary of different criteria (Av-SL, NMW and JE) mentioned in the main text, with examples of maps according to this characterization.

window ( $E_i \in T_{N,\delta}$ ), the final energy  $E_f$  is also within this energy window except with probability  $2p(T_{N,\delta}^c)$ , which is (sub-)exponentially small in  $N$ . We will use this later.

We are now ready to evaluate the probability of obtaining macroscopic work.

$$\begin{aligned} p(w \geq a) &= p(T_{N,\delta}) \cdot p(w \geq a | E_i \in T_{N,\delta}) \\ &\quad + p(T_{N,\delta}^c) \cdot p(w \geq a | E_i \in T_{N,\delta}^c) \\ &\leq p(w \geq a | E_i \in T_{N,\delta}) + p(T_{N,\delta}^c). \end{aligned}$$

We can estimate the first term as

$$p(w \geq a | E_i \in T_{N,\delta}) \leq p(E_f \leq (e^{(N)} + \delta - a)N | E_i \in T_{N,\delta}).$$

We now choose  $\delta = a/2$  and get

$$\begin{aligned} p(w \geq a | E_i \in T_{N,\delta}) &\leq p(E_f \leq (e^{(N)} - a/2)N | E_i \in T_{N,\delta}) \\ &\leq \text{Tr} \left[ \sigma'_2 \left( \mathbf{I} - \Pi_{a/2}^{(N)} \right) \right] \\ &\leq 2p(T_{N,a/2}^c), \end{aligned}$$

where we have used (10) in the last step. Altogether, we thus find

$$p(w \geq a) \leq 3p(T_{N,a/2}^c),$$

which decays to zero (sub-)exponentially by (8). This concludes the proof.  $\square$

As a side-remark, we note that if the Gibbs-preserving channels that appear here are interpreted as modelling the interaction with a heat bath, then the above result can be interpreted as a "no macroscopic heat" statement: If a macroscopic system is brought in thermal contact with a heat bath at the same temperature, then the probability of an exchange of a macroscopic amount of heat is arbitrarily small in the system size.

## B Microscopic toy example

In this section, we show that already for small systems and using catalysts, the JE can be violated. We do so by constructing non-unital catalytic channels. Indeed, such maps can be realized “quasi-classically”, in the sense that in the construction it is sufficient to consider the energy spectra of the involved states and that all unitaries are simple permutations of those values. We consider a 3-level system with energy levels  $E_1 = 0, E_2 = 0, E_3 = \Delta$  in the thermal state

$$w = \left( \frac{1}{Z}, \frac{1}{Z}, \frac{Z-2}{Z} \right),$$

where  $Z = 2 + \exp(-\beta\Delta)$  is the partition function and we express the state as a probability vector, such that  $w_i$  denotes the  $i$ th eigenvalue of the thermal state. For later, we observe that  $2 \leq Z \leq 3$ .

We are going to construct a simple non-unital catalytic channel that involves a 2-dimensional catalyst. Let  $e_i$  and  $f_j$  denote the basis states for the vector spaces  $\mathcal{V}_S$  and  $\mathcal{V}_C$  describing the system and catalyst respectively. We define the permutation  $\pi$  acting on the joint vector space  $\mathcal{V}_S \otimes \mathcal{V}_C$  as that permutation which exchanges the respective levels  $e_1 \otimes f_1 \Leftrightarrow e_2 \otimes f_2$  and  $e_2 \otimes f_1 \Leftrightarrow e_3 \otimes f_2$  and leaves all other entries unchanged (see Fig. 3). For the catalyst to remain unchanged for this permutation and initial system state, it is easy to check that the catalyst has to be given by the vector

$$q = \left( \frac{Z-1}{Z+1}, \frac{2}{Z+1} \right).$$

Now, the catalytic channel  $\mathcal{C}$  induced by this catalyst and permutation on the system has the general effect

$$\mathcal{C}(p_1, p_2, p_3) = (q_1 p_2 + q_1 p_3, q_1 p_1 + q_2 p_3, q_2 p_1 + q_2 p_2),$$

so that, in particular, the maximally mixed input state is mapped to

$$\mathcal{C}(\mathbf{I}/3) = \frac{2}{3} \left( \frac{Z-1}{Z+1}, \frac{1}{2}, \frac{2}{Z+1} \right),$$

which is different from the maximally mixed vector for any  $\Delta > 0$ .

What is more, we can also directly calculate the work-distribution  $p(w)$ , yielding

$$\begin{aligned} p(0) &= \frac{1}{Z(Z+1)} [Z + 3 + 2(Z-2)(Z-1)], \\ p(\Delta) &= \frac{2(Z-2)}{Z(Z+1)}, \\ p(-\Delta) &= \frac{Z-1}{Z(Z+1)}. \end{aligned}$$

We now want to compute  $\langle e^{\beta W} \rangle$ . To do so, it is useful to note that  $e^{-\beta\Delta} = Z-2$  and hence  $e^{\beta\Delta} = 1/(Z-2)$ . We find

$$\langle e^{\beta W} \rangle = \frac{Z+5+2(Z-2)(Z-1)}{Z(Z+1)} \geq 1.$$

$q_2 p_3$	$q_1 p_3$	$p_3$	$\rightarrow$	$q_1 p_2$	$q_1 p_3$	$q_1 p_2 + q_1 p_3$
$q_2 p_2$	$q_1 p_2$	$p_2$		$q_1 p_1$	$q_2 p_3$	$q_1 p_1 + q_2 p_3$
$q_2 p_1$	$q_1 p_1$	$p_1$		$q_2 p_1$	$q_2 p_2$	$q_2 p_1 + q_2 p_2$
$q_2$	$q_1$			$q_2$	$q_1$	

Figure 3: We represent the joint state of system and catalyst by means of a table. *Left*: At the beginning the joint system starts out in a product state, so that the entry  $(i, j)$  is given by the product of the  $i$ th eigenvalue of the system and  $j$ th eigenvalue of the catalyst. *Right*: After applying the permutation highlighted in red, the marginal state of the system, given by the rows sums, has changed, while the marginal state of the catalyst (given by the column sums), has to remain invariant. For a two-dimensional catalyst, specifying the permutation and initial system state fixes the catalyst state.

In fact, this quantity is larger than 1 whenever  $Z < 3$ , corresponding to  $\Delta > 0$ . Its maximum is given as  $7/6$  for  $Z = 3$ , which corresponds to  $\Delta \rightarrow \infty$ . Thus, the Jarzynski inequality is violated. At the same time the second law is fulfilled as expected, since  $p(-\Delta) \geq p(\Delta)$ .

## C Work extraction for initial micro-canonical ensembles

In this appendix, we show that a statement similar to Proposition 1 of the main text holds in the slightly different setting of a micro-canonical initial state. This serves two purposes: i) in statistical mechanics, one often assumes that closed, macroscopic systems are described by micro-canonical ensembles due to the postulate of equal a priori probabilities of microstates corresponding to a macrostate. ii) The proof for the microcanonical initial state is conceptually simpler, but also provides the blueprint for the slightly more involved proof in the case of a canonical state, which is provided in Sec. D.

In the following, we denote by  $I \subset \mathbb{R}$  an energy window, by  $g(I)$  the number of energy eigenstates in this window,

$$g(I) = \sum_{E_i \in I} 1,$$

and the corresponding micro-canonical state by

$$\Omega_S(I) = \frac{1}{g(I)} \sum_{E_i \in I} |E_i\rangle\langle E_i|.$$

A micro-canonical energy window around energy density  $e$  is any energy window  $I(e)$  of the form  $[e - O(\sqrt{N}), e]$ , where  $N$  is the number of particles.

The only difference to the standard setting described in the main text (as depicted in Fig. 1) is that the initial state differs from the thermal state  $\omega_\beta(H)$ . Instead, it is given by the micro-canonical ensemble. In other words, given a micro-canonical energy window  $I$ , we consider channels

$\mathcal{C}$  of the form

$$\begin{aligned} \mathcal{C}(\cdot) &= \text{Tr}_C(U(\cdot \otimes \sigma_C)U^\dagger) \\ \text{s.t. } \text{Tr}_S(U(\Omega_S(I) \otimes \sigma_C)U^\dagger) &= \sigma_C. \end{aligned}$$

We carry over notation from the main text, so that  $p(w \geq \epsilon)$  denotes the probability of measuring the system's energy per particle decrease by at least an amount  $\epsilon$ , and so on. Furthermore, we take the catalyst Hamiltonian in our construction to be  $H_C = \mathbf{I}$ .

We will now first show that the NMW principle also holds for micro-canonical states of generic many-body systems. After that we will show that it can be circumvented using catalysts. To show the validity of the NMW principle we will use the same reasoning as presented in Ref. [35], where the NMW principle has been studied before. Thus, the following proof is essentially a reproduction for the convenience of the reader. We consider a sequence of many-body Hamiltonians  $H_S^{(N)}$  on  $N$  particles with the generic property of having an exponential density of states:

$$g((-\infty, E]) := \sum_{E_i \leq E} 1 = e^{N\mu(E/N) - o(N)}, \quad (11)$$

where  $\mu$  is a strictly monotonic and differentiable function independent of  $N$  and  $o(N)$  denotes terms small compared to  $N$ ,  $\lim_{N \rightarrow \infty} o(N)/N = 0$ .

**Proposition 4** (NMW for micro-canonical states). *Consider a sequence of  $N$ -particle Hamiltonians fulfilling (11) and a sequence of micro-canonical energy windows  $I^{(N)} = [eN, eN + \delta\sqrt{N}]$  around energy density  $e$  (with  $\delta > 0$  fixed). Then for any unital channel acting on the  $N$ -particle system, the probability of extracting work  $w$  per particle is bounded as*

$$p(w > \epsilon) \leq C e^{-\mu'(\epsilon)\epsilon N + o(N)},$$

where  $C > 0$  is a constant and  $\mu'$  denotes the derivative of  $\mu$ .

*Proof.* Let  $I_{\leq} := (-\infty, (e - \epsilon)N + \delta\sqrt{N}]$ , denote by  $P_S(I_{\leq})$  the projector onto energy-eigenstates with energies below  $(e - \epsilon)N + \delta\sqrt{N}$  and let  $\mathcal{U}$  denote a unital channel. In the following, we write  $I$  instead of  $I^{(N)}$  to simplify notation. Then

$$\begin{aligned} p(w > \epsilon) &\leq \text{Tr}(P_S(I_{\leq})\mathcal{U}[\Omega_S(I)]) \\ &= \sum_{E_i \in I} \frac{1}{g(I)} \text{Tr}(P_S(I_{\leq})\mathcal{U}[|E_i\rangle\langle E_i|]) \\ &\leq \frac{1}{g(I)} \text{Tr}(P_S(I_{\leq})\mathcal{U}[\mathbf{I}]) = \frac{g(I_{\leq})}{g(I)}. \end{aligned}$$

Writing  $\bar{e} := e + \delta N^{-1/2}$ , we have

$$\begin{aligned} g(I) &= e^{N\mu(\bar{e}) - o(N)} - e^{N\mu(e) - o(N)} \\ &= e^{N\mu(\bar{e}) - o(N)} \left(1 - e^{-N(\mu(\bar{e}) - \mu(e)) + o(N)}\right) \\ &\approx e^{N\mu(\bar{e}) - o(N)}, \end{aligned}$$

where in the last estimation we use that  $\mu$  is strictly monotonic. In particular, we can estimate the exponential in the parenthesis as

$$e^{-N(\mu(\bar{e}) - \mu(e)) - o(N)} = O\left(e^{-\delta\mu'(\epsilon)N^{1/2}}\right),$$

where  $\mu'$  denotes the derivative of  $\mu$ . Using  $g(I_{\leq}) = e^{N(\mu(\bar{e}) - \mu(e)) - o(N)}$  we then find

$$\begin{aligned} p(w > \epsilon) &\leq \frac{e^{-N(\mu(\bar{e}) - \mu(e)) + o(N)}}{1 - O(e^{-\delta\mu'(\epsilon)\sqrt{N}})} \\ &\leq C e^{-\mu'(\epsilon)\epsilon N}. \end{aligned}$$

□

We have here used that  $\mu$  is differentiable to prove this result. Similar results would follow for weaker notions of regularity of  $\mu$ , such as Lipschitz-continuity. Having proven the NMW principle for generic many-body systems, let us now show how to circumvent it using catalysts.

**Proposition 5** (Overcoming NMW using catalysts). *Consider a Hamiltonian  $H_S$  and a microcanonical state  $\Omega_S(I)$ , with  $I$  a micro-canonical energy window around energy density  $e$ . Suppose there exists an energy window  $I_+$  with  $g(I_+) = g(I)^2$ . Then, for any  $0 \leq e_- < e$ , there exists a catalytic channel such that*

$$p(w \geq e - e_-) = \frac{1}{2}.$$

Before giving the proof of the proposition, let us emphasize again that the required conditions on the Hamiltonian are very weak. In particular, the conditions are (approximately) fulfilled if the density of states is well approximated by an exponential in the range of energies that we are working in, a condition that is typically fulfilled in many-body systems and, as we have seen above, leads to an NMW principle if we do not allow for catalysts.

*Proof.* A sketch of the proof is given in Fig. 4. The proof is constructive in the sense that we provide an explicit catalyst and unitary. We first introduce some useful notation. Define  $g := g(I)$ ,  $g_+ := g(I_+) = g^2$  and let  $P_S(I)$  and  $P_S(I_+)$  be the projectors onto the corresponding energy subspaces. Let  $|E_- \rangle$  be any eigenstate of the Hamiltonian such that  $0 \leq E_-/N = e_- \leq e$ . Following this notation, the initial state of the system is

$$\Omega_S(I) = \frac{1}{g} P_S(I).$$

The aim is to bring the system to a state that is an equal mixture of  $|E_- \rangle \langle E_-|$  and  $\Omega(I_+)$ . To do this, we employ a catalyst of dimension  $d_C = g + 1$ . Let  $\{|i\rangle_C\}_{i=1}^{d_C}$  be an arbitrary orthonormal basis on the Hilbert-space of the catalyst and let  $P_C = \sum_{i=1}^g |i\rangle\langle i|$ . The initial state on the catalyst is given by

$$\sigma = \frac{1}{2g} P_C + \frac{1}{2} |d_C\rangle\langle d_C|_C.$$

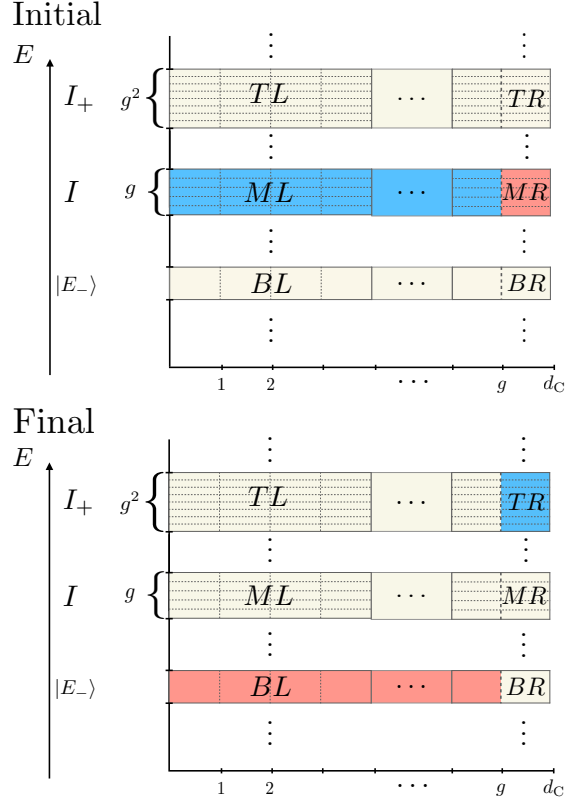


Figure 4: Proof sketch for Proposition 5: *Top*: We represent the initial product state of system and catalyst by means of a table, using the fact that both are initially diagonal in the energy eigenbasis: Ordering the spectra of both states non-increasingly, the entry  $(i, j)$  of the table corresponds to the product of the  $i$ -th eigenvalue of the system (corresponding to the a particular energy eigenstate) and the  $j$ -th energy eigenvalue of the catalyst. We focus on three regions in the table—denoted top (T), middle (M), bottom (B)—corresponding to two degeneracy bands  $I$  and  $I_+$  and (the projector onto) a single energy eigenvector  $|E_-\rangle$ : Since the system is initially in the micro-canonical ensemble with energy window  $I$ , the support of the joint state is initially contained in the coloured middle band. The catalyst is constructed as carrying half of its mass uniformly distributed over  $d_C - 1$  of its entries and the other half in a single entry. This means that the middle band is divided into two subregions, middle left (ML) and middle right (MR), where the total probability mass coloured in blue equals the mass coloured in red. Furthermore, each of these subregions has its mass uniformly distributed over its entries. *Bottom*: By construction, both the subregions BL and MR as well as ML and TR have the same number of entries. Hence, we can swap BL and MR by means of a permutation, and similarly for ML and TR. This permutation results in a reduced state on  $S$  of the form Eq. (13) and hence produces the claimed work extraction probability. Moreover, it leaves the marginal state of the catalyst unchanged, so that the permutation induces a valid catalytic channel.

We define the unitary  $U$  by the conditions

$$\begin{aligned} U[P_S(I) \otimes |d_C\rangle\langle d_C|_C]U^\dagger &= |E_-\rangle\langle E_-| \otimes P_C \\ U[P_S(I) \otimes P_C]U^\dagger &= P_S(I_+) \otimes |d_C\rangle\langle d_C|_C. \end{aligned}$$

This is possible since i) the corresponding subspaces have the same dimension, ii) the subspaces in the

two equations are orthogonal and iii) subspaces of the same dimension can always be mapped into each other by a unitary. In fact there will be many different unitaries achieving this, and any of them is fine for our purposes.

Applying  $U$  to the state  $\Omega_S(I) \otimes \sigma_C$  one obtains

$$\begin{aligned} U(\Omega_S(I) \otimes \sigma_C)U^\dagger &= \frac{1}{2g^2}U(P_S(I) \otimes P_C)U^\dagger + \frac{1}{2g}U\left(P_S(I) \otimes \frac{1}{2}|d_C\rangle\langle d_C|_C\right)U^\dagger \\ &= \frac{1}{2}\Omega_S(I_+) \otimes |d_C\rangle\langle d_C|_C + \frac{1}{2g}|E_-\rangle\langle E_-| \otimes P_C. \end{aligned} \quad (12)$$

It is clear from (12) that

$$\text{Tr}_S(U(\Omega_S(I) \otimes \sigma_C)U^\dagger) = \sigma_C,$$

as required for a catalytic channel. Moreover, the quantity of interest  $P(w \geq e - e_-)$  given by this channel  $\mathcal{C}$  (defined by  $U$  and  $\sigma_C$ ) can be derived by noting that

$$\mathcal{C}(\Omega_S(I)) = \frac{1}{2}\Omega(I_+) + \frac{1}{2}|E_-\rangle\langle E_-|, \quad (13)$$

so that  $p(W \geq e - E'/n) = \frac{1}{2}$ .  $\square$

## D Proof of Proposition 1 in the main text

In this section, we provide the proof and full statement of Proposition 1 in the main text. This proof is very similar to that of the micro-canonical case presented in the previous section, we will hence only describe the adjustments that have to be made. Also, unlike in Appendix C, we now again consider the standard setting and definition of catalytic channels as introduced in the main text. In the following, we denote by  $P_S(I)$  the projector onto a specific energy-window  $I$ . Then  $g(I)$  is equal to the rank of  $P_S(I)$ . We consider Hamiltonians  $H_S^{(N)}$  on a regular lattice  $\Lambda^{(N)}$  of  $N$  sites and assume that the  $H_S^{(N)}$  (for different values of  $N$ ) constitute a sequence of *local, uniformly bounded* Hamiltonians:

$$H_S^{(N)} = \sum_{x \in \Lambda^{(N)}} h_x,$$

where each term  $h_x$  acts on sites at most a distance  $l$  away from  $x$  and the norm of each term is bounded as  $\|h_x\| \leq h$  independent of the system size for some constant  $h$ .

**Proposition 6** (Lower bound to the probability of work extraction). *Fix an inverse temperature  $\beta > 0$  and consider a sequence of local, uniformly bounded  $N$ -particle Hamiltonians  $H_S^{(N)}$  on a regular,  $D$ -dimensional lattice. Assume that the states  $\omega_\beta(H_S^{(N)})$*

*have a finite correlation length bounded by a constant and denote by  $e^{(N)}$  the energy density corresponding to  $\beta$ . Let  $\delta > 0$  be fixed and consider  $I^{(N)} := [e^{(N)}N - \delta\sqrt{N}, e^{(N)}N]$ . Further assume that there exist micro-canonical energy windows  $I_+^{(N)}$  with  $g(I_+^{(N)}) = g(I^{(N)})^2$ . Then, for sufficiently large  $N$ , there exists, for any  $0 < e_- < e^{(N)}$ , a corresponding sequence of catalytic channels such that*

$$p(w \geq e^{(N)} - e_-) \geq 1/2 - Ce^{-\frac{(\delta^2 N)^{\frac{1}{1+D}}}{K}},$$

where  $C, K > 0$  are constants.

Before giving the proof, we again emphasize the weakness of the assumptions in the statement, which, in the limit of large  $N$ , can be satisfied to arbitrary precision if the density of states grows exponentially within  $I^{(N)}$ , as is typically the case. Furthermore, let us emphasize that the energy densities  $e^{(N)}$  fluctuate arbitrarily little (for sufficiently large  $N$ ) from a constant  $e$  due to the locality of temperature [36].

*Proof.* The proof follows the proof for the micro-canonical case in Appendix C. In particular, the unitary that we use is exactly the same as that constructed in the proof for the micro-canonical case. However, here we do not construct the state of the catalyst explicitly, but allude to Lemma 4, which ensures there is always some catalyst given the unitary that we consider. What remains to be done is to show that for every such catalyst the probability distribution of work is as claimed. To do this, we denote by  $r$  the initial probability of an energy-window  $I$  in the initial thermal state given by

$$r(I) = \text{Tr}(P_S(I)\omega_\beta(H_S))$$

and by  $r_- = \langle E_- | \omega_\beta(H_S) | E_- \rangle$  the initial weight on the low-energy eigenstate  $|E_- \rangle$ . Here and in the following, we drop the explicit dependence on the system-size for simplicity of notation. The following arguments work as long as  $N$  is large enough such that



the energy-windows  $I$  and  $I_+$  are disjoint. Denote by  $\{q_i\}_{i=1}^{d_C}$  the spectrum of the catalyst. By considering the action of the used unitary, it is easy to see that a necessary condition for the transition being catalytic under the given unitary is that

$$q_{d_C} (r(I) + r(I_+)) = (1 - q_{d_C})(r(I) + r_-). \quad (14)$$

This can be seen, for example, from Fig. 4, where the above represents the condition of catalyticity for the right-most column. Solving in (14) for  $q_{d_C}$ , we find that

$$q_{d_C} = \frac{r(I) + r_-}{2r(I) + r_- + r(I_+)}.$$

We now invoke the result from Ref. [33] (as previously in the proof of Lemma 3) which implies that

$$r(I) \geq 1 - \epsilon_N,$$

where there exist constants  $C, K$  such that

$$\epsilon_N \leq C e^{-\frac{(\delta^2 N)^{\frac{1}{1+D}}}{K}}.$$

For large enough  $N$ , the energy windows  $I$  and  $I_+$  are disjoint. Hence  $0 \leq r_- + r(I_+) \leq 1 - r(I)$  and we find

$$\begin{aligned} q_{d_C} &\geq \frac{r(I)}{2r(I) + 1 - r(I)} = \frac{r(I)}{1 + r(I)} \\ &\geq \frac{r(I)}{2} \geq \frac{1}{2} (1 - \epsilon_N). \end{aligned}$$

Finally, we find

$$\begin{aligned} p(w \geq e - e_-) &\geq P(E_f = E_- | E_i \in I) w(I) = q_{d_C} \cdot r(I) \\ &\geq \frac{1}{2} (1 - \epsilon_N)^2 \geq \frac{1}{2} - \epsilon_N. \end{aligned}$$

□

**Lemma 4** (Existence of catalysts). *Let  $\rho_S$  be a quantum state on a finite-dimensional Hilbert-space  $\mathcal{H}_S$  and  $U$  be a unitary on the Hilbert-space  $\mathcal{H}_S \otimes \mathcal{H}_C$ , where  $\mathcal{H}_C$  is an arbitrary finite-dimensional Hilbert-space. Then there exists a density matrix  $\sigma_C$  such that*

$$\text{Tr}_S (U(\rho_S \otimes \sigma_C)U^\dagger) = \sigma_C.$$

*Proof.* The map  $\sigma_C \mapsto \text{Tr}_S (U(\rho_S \otimes \sigma_C)U^\dagger)$  specifies a quantum-channel. Since every quantum channel is a continuous map on the compact and convex set of states, it has a fixed point by Brouwer's fixed point theorem ([37], Section 4.2.2). □

## E Proof of Proposition 3 in the main text

Proposition 3 in the main text follows straightforwardly once we realize that we can tune the process used in the

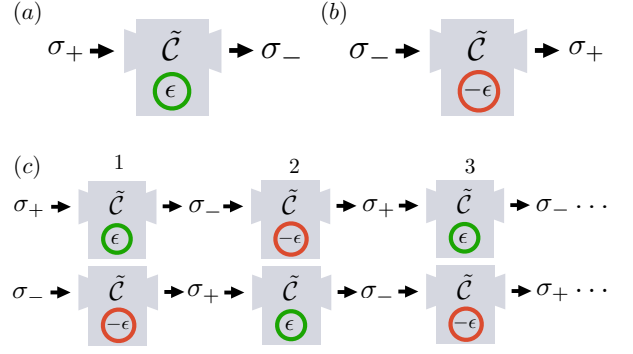


Figure 5: The idea behind the proof of Proposition 3 in the main text: For any choice of unitary, we can understand the second condition in the Def. 2 of the main text as the definition of a quantum channel  $\tilde{C}$  acting on  $C$ . We find a  $\tilde{C}$  and two states  $\sigma_-, \sigma_+$  with the following properties: (a) If the initial state of the catalyst is  $\sigma_+$ , the result of running the standard protocol is to extract positive work  $\epsilon$  from the system, while the state of the catalyst is changed to  $\tilde{C}(\sigma_+) = \sigma_-$ . (b) The *same* unitary, however, for initial state  $\sigma_-$ , extracts negative work  $-\epsilon$  and changes the catalyst state to  $\tilde{C}(\sigma_-) = \sigma_+$ . (c) Hence, if we initialize the catalyst in the state  $\sigma = \frac{1}{2}(\sigma_+ + \sigma_-)$ , then there are two “branches” of work extraction distributions, each occurring with probability  $1/2$ , while the resulting channel on  $S_i$  is catalytic for every  $i$ . Note that, if the agent *knew* whether her input state was  $\sigma_+$  or  $\sigma_-$ , then she could condition her unitary  $U$  on this knowledge and achieve the claimed work distribution easily. Hence, the key achievement of the proof is to show that agents can achieve correlated work distributions *without knowing the initial state of the catalyst*.

construction of the proof for Proposition 1 in the main text in such a way that its repeated application implies the claimed work distribution. This follows because we have great freedom in choosing the state  $E_-$ . In particular, in terms of notation of the previous section, let  $e_+^{(N)}$  denote the energy density around which the window  $I_+^{(N)}$  is centered. Then we choose  $E_-$  in such a way that  $e - E_-/N = e_+ - e$  to ensure that the extracted and invested amount of work in every iteration are exactly the same. The above choice of  $E_-$  is always possible for the Hamiltonians with exponentially growing density of states that we consider (for which  $e_+$  will not be much greater than  $e$ ).

Fig. 5 provides a sketch of the proof. For the many-player process described in the main text, let

$$p(w_2, w_3, w_4, \dots | w_1)$$

denote the work probability distribution for agents 2 to  $n$  conditional on the player 1 extracting work  $w_1$ . The key recognition then is that, for any  $n$ , by construction of the catalytic channel,

$$p(w_2, w_3, \dots | w_1) = 1 \quad (15)$$

whenever  $w_i = -w_{i-1}$  for all  $i \in \{2, \dots, n\}$ , while

$$p(w_2, w_3, \dots | w_1) = 0$$

in all other cases. This is because, if the extracted work in the first round was negative, corresponding to an increase in the system's energy, then by construction of the unitary, the final state of the catalyst is  $\sigma' = |d\rangle\langle d|$  with probability one, since all transitions that lead to an increase in energy on the system result in this final state. This, in turn, is sufficient to determine that, for the second player, the application of the unitary to this catalyst state  $\sigma'$  and her copy of  $\omega_\beta(H)$  will result in a decrease of the system's energy (and hence positive work extraction) and a final catalyst state  $\sigma''$  with support on the subspace  $\sum_i^g |i\rangle\langle i|$ , etc. This reasoning can be applied to an arbitrary number of agents and also to the case in which the extracted work in the first round was positive, and hence implies (15). The claimed work distributions then follow from

$$p(w_1, w_2, \dots, w_n) = p(w_2, w_3, w_4, \dots | w_1) p(w_1),$$

together with Proposition 1 in the main text. We also note that a similar conclusion holds in the case of the microscopic toy-example presented in Section B, where this behaviour can be checked easily by explicit calculation.

## F Proof of Proposition 2 in the main text

Given a catalytic channel  $\mathcal{C}$ , let  $\mathcal{U}$  denote the unitary channel applied to the joint system  $SC$  when dilating the channel. The key observation is that, if  $\mathcal{U}$  is unitary, then  $\mathcal{U}^*$  is trace-preserving and hence maps quantum states to quantum states (in fact, this property holds for all unital channels). Here,  $*$  denotes the Hilbert-Schmidt adjoint. We then write

$$\begin{aligned} \langle e^{\beta W} \rangle &= \text{Tr}(\omega \mathcal{C}(\mathbf{1})) \\ &= \text{Tr}(\omega \otimes \mathbf{I} \mathcal{U}(\mathbf{1} \otimes \sigma)) \\ &= d_C \text{Tr} \left( \mathcal{U}^* \left( \omega \otimes \frac{\mathbf{1}}{d_C} \right) \mathbf{1} \otimes \sigma \right) \\ &\leq d_C \|\mathbf{1} \otimes \sigma\|_\infty \left\| \frac{\mathbf{1}}{d_C} \otimes \sigma \right\|_1 \\ &= d_C \|\sigma\|_\infty. \end{aligned}$$

Here, the first equality is simply Eqn. 2 in the main text and we write  $\omega$  instead of  $\omega_\beta(H)$ . Similarly, we get

$$\begin{aligned} \langle e^{\beta W} \rangle &= d \text{Tr} \left( (\omega \otimes \mathbf{I}) \mathcal{U} \left( \frac{\mathbf{1}}{d} \otimes \sigma \right) \right) \\ &\leq d \|\omega \otimes \mathbf{I}\|_\infty = d \|\omega\|_\infty. \end{aligned}$$

## G Non-trivial Hamiltonian on the catalyst

In this section we show that the probability distribution of work done on the system is independent of the Hamiltonian on the catalyst. To do this, let us first assume we had a catalytic process that uses a catalyst with a non-trivial Hamiltonian  $H_C$  and a quasi-classical state  $\sigma_C$ , i.e.,  $[H_C, \sigma_C] = 0$ . We assume that  $\sigma_C$  is quasi-classical, since it is well known that it is impossible to associate a meaningful random variable of work in the case coherent initial states [38]. Using the two-time measurement process on the system and catalyst together, we can then associate a bi-partite work-distribution  $P(W^{(S)}, W^{(C)})$ , where

$$W^{(S)} = E_f^{(S)} - E_i^{(S)}$$

denotes the work done on the system and

$$W^{(C)} = E_f^{(C)} - E_i^{(C)}$$

the work done on the catalyst. The work distribution on the system is simply given by the marginal

$$P(W^{(S)}) = \int P(W^{(S)}, W^{(C)}) dW^{(C)}.$$

Let us write  $\sigma_C = \sum_j \sigma_j |E_j^{(C)}\rangle\langle E_j^{(C)}|$  and  $\omega_\beta(H) = \sum_k w_k |E_k^{(S)}\rangle\langle E_k^{(S)}|$ . We then get

$$\begin{aligned} P(W^{(S)}) &= \sum_{E_f^{(S)} - E_i^{(S)} = W^{(S)}} \sum_{E_{f'}^{(C)}} \sum_{E_{i'}^{(C)}} P(E_f^{(S)}, E_{f'}^{(C)} | E_i^{(S)}, E_{i'}^{(C)}) P(E_i^{(S)}) P(E_{i'}^{(C)}) \\ &= \sum_{E_f^{(S)} - E_i^{(S)} = W^{(S)}} \sum_{E_{f'}^{(C)}} \sum_{E_{i'}^{(C)}} \langle E_f^{(S)} | \otimes \langle E_{f'}^{(C)} | \left( U \left( w_i \sigma_{i'} |E_i^{(S)}\rangle\langle E_i^{(S)}| \otimes |E_{i'}^{(C)}\rangle\langle E_{i'}^{(C)}| \right) U^\dagger \right) |E_f^{(S)}\rangle \otimes |E_{f'}^{(C)}\rangle \\ &= \sum_{E_f^{(S)} - E_i^{(S)} = W^{(S)}} \langle E_f^{(S)} | \text{Tr}_C \left( U \left( w_i |E_i^{(S)}\rangle\langle E_i^{(S)}| \otimes \sigma \right) U^\dagger \right) |E_f^{(S)}\rangle \\ &= \sum_{E_f^{(S)} - E_i^{(S)} = W^{(S)}} \langle E_f^{(S)} | \mathcal{C} \left( w_i |E_i^{(S)}\rangle\langle E_i^{(S)}| \right) |E_f^{(S)}\rangle. \end{aligned}$$



It is hence identical with the one obtained on the system alone when we think of the catalyst as a system with a trivial Hamiltonian, that is, with the distribution as defined above Eqn. 2 in the main text. This shows that we can always assume that the catalyst has a trivial Hamiltonian, in which case it is clear that no energy flows from the catalyst to the system, even probabilistically. Therefore, such an energy flow is not necessary to implement catalytic transitions.

## H Comparison with literature on generalized Jarzynski equalities in the presence of correlations

In recent years, the role of correlations, specifically quantified by the mutual information, has been well studied, in particular with respect to its influence on the Jarzynski equality [25, 26], even leading up to experimental demonstrations to test these theoretical results [28, 39]. One may ask, how the results of this manuscript fit in the context of that line of research. This section provides a brief overview of the main differences.

In Ref. [25] the core observation is that the presence of initial correlations between a system  $S$  and an ancillary (catalyst)  $C$  can be used to create a thermodynamic advantage, in the sense that such processes obey a generalized JE and Second law and hence can be used to by-pass the constraints imposed by the original JE and Second law. Specifically, [25] derives (according to their generalized version of Jarzynski equality) a bound on the work *performed* on the system that is given by

$$\langle W \rangle \geq \langle \Delta F \rangle + \langle \Delta E_{\text{int}} \rangle + \beta^{-1} \langle \Delta I \rangle,$$

where  $\langle \Delta F \rangle$  is the difference between final and initial equilibrium free energy on the system,  $\langle \Delta E_{\text{int}} \rangle$  for the energy difference coming from the interaction Hamiltonian between system and catalyst, and finally  $\langle \Delta I \rangle$  is the change in mutual information between system and catalyst. For our setup, both  $\langle \Delta F \rangle$  and  $\langle \Delta E_{\text{int}} \rangle$  are zero. Given that the extracted work  $W_{\text{ext}} = -W$ , the above bound reduces to

$$\langle W_{\text{ext}} \rangle \leq -\beta^{-1} \langle \Delta I \rangle,$$

which says that if one allows the consumption of mutual information (leading to  $\Delta I < 0$ ), then it is possible to violate the average second law, namely extract some positive amount of  $W_{\text{ext}}$  from a Gibbs state, for instance by reducing the entropy of the system in the process. This particular viewpoint of correlations (information) being a thermodynamic resource is a mature and well-studied one.

In our setting, however, the initial state of system and catalyst are always uncorrelated, which means that we always have  $\langle \Delta I \rangle \geq 0$ . Hence it is clear that the type of catalytic operation studied in Ref. [25] cannot correspond to our setting, since the generalized JE and Second law allow for violations of the original JE and Second law *only* if

$\langle \Delta I \rangle < 0$ . The difference to our setting, however, is easily understood. It lies in the fact that here we allow for more general joint evolutions of the system and the catalyst. Indeed, it is easy to see that under the requirement that the initial state between catalyst and system be uncorrelated, the channels that can be implemented on the system via the operations allowed in Ref. [25] are unital channels, for which we show above that they cannot be used to by-pass the JE (see Fig. 2). This is because in the above works, the catalyst is required to not evolve over time. In contrast, the notion of a  $\beta$ -catalytic channel allows for the evolution of the catalyst to be non-trivial, as long as the final density matrix describing the catalyst is unchanged. Since this constraint only requires the *statistical* invariance of the catalyst, this allows for a much broader class of evolutions to be implemented on the system and hence explains how we can by-pass the JE and NMW in a setting where the marginal entropy of the system has to increase. In summary, the key differences to the line of work rooted in Refs. [25, 26] are that we study processes that by-pass the JE by means of the *creation* of correlations paired with catalysts that evolve non-trivially over time, while in the above work processes are studied that by-pass the JE by means of the *absorption* of initial correlations paired with catalysts that do not evolve over time.

## I Is it necessary to establish correlations with the catalyst?

In our definition of  $\beta$ -catalytic channels, we allow the catalyst to become correlated with the system. These correlations are certainly necessary for the correlated multiplayer strategies discussed in the main text, but one might wonder whether they are also necessary to violate NMW on a single system. To make this question concrete, consider the set of  $\beta$ -trumping channels, where a quantum channel  $\mathcal{T}$  is in this set iff it has the form

$$\mathcal{T}(\rho) = \text{Tr}_2(\mathcal{N}(\rho \otimes \sigma)),$$

where  $\mathcal{N}(\mathbb{I}) = \mathbb{I}$  is unital and  $\mathcal{N}(\omega_\beta(H) \otimes \sigma) = \rho' \otimes \sigma$ . Note that in the case of  $\beta$ -catalytic channels, we restricted the corresponding channel  $\mathcal{N}$  to be unitary. Here, we allow instead for the more general class of unital channels. We will prove that in the unitary case, NMW cannot be violated by  $\beta$ -trumping channels even though Jarzynski's equality may be violated. We will also present arguments that suggest that the same is true in the unital case.

It is worth noting, for starters, that the fully thermalizing channel is exactly a  $\beta$ -trumping channel where  $\sigma = \omega_\beta(H)$ , and  $\mathcal{N}$  is a unitary swap between the system and catalyst. Thus, even in the case of a unitary channel  $\mathcal{N}$ , such  $\beta$ -trumping channels can violate Jarzynski's equality. On the other hand, in the main text we have demonstrated that the thermalizing channel cannot violate NMW since it is Gibbs preserving. Hence, the above leaves open the question whether the NMW condition can be violated by

means of  $\beta$ -trumping channels. However, we do not believe that this is the case, for the following reasons:

i) Our constructions of violating NMW can *not* work in the trumping case. This is because in the trumping setting the so-called min-entropy  $S_\infty$  (minus log of the largest eigenvalue) of the final state has to be at least as large as that of the initial state (see for example Ref. [6]). However, in our constructions, the final min-entropy is essentially given by  $-\log(p(w \geq \epsilon)) \approx \log(2)$ , whereas the initial min-entropy is extensive in  $N$ . It thus *decreases* by a macroscopic amount.

ii) The previous point also suggests a route for arguing that  $\beta$ -trumping channels cannot be used to violate NMW in general: We now present an argument that rules out violations of NMW in the case of a microcanonical initial state  $\Omega$  with energy density  $e$ , but we expect that similar statements hold true for the canonical case due to equivalence of ensembles-type of arguments. Because of the highly peaked probability distribution of the energy density for a macroscopic, non-critical many-body system, it is easy to see that the probability  $p(w \geq \epsilon)$  to extract work per particle at least  $\epsilon$  is (up to arbitrarily small corrections for large  $N$ ) given by the total probability of measuring an energy below  $(e - \epsilon)N$  in the final state  $\mathcal{T}(\Omega)$ . Let us denote the projector onto these energies by  $P$ . We then have

$$p(w \geq \epsilon) \approx \text{Tr}[P\mathcal{T}(\Omega)],$$

where the approximation is arbitrarily good as  $N \rightarrow \infty$ . This insight also was an essential ingredient to the proof that Gibbs-preserving maps cannot violate NMW. Now, to leading order, the total number of states with energy below  $(e - \epsilon)N$  is given by  $\exp(s(e - \epsilon)N)$ , where  $s(e - \epsilon)$  is the microcanonical entropy density at energy density  $e - \epsilon$ . Since the total weight in this subspace is  $p(w \geq \epsilon)$ , the final min-entropy is upper bounded by

$$S_{\min}^{(\text{final})} \leq -\log(p(w \geq \epsilon)) + s(e - \epsilon)N.$$

However, since trumping requires  $S_{\min}^{(\text{final})} \geq S_{\min}^{(\text{initial})} = s(e)N$ , we then find

$$p(w \geq \epsilon) \leq \exp(-(s(e) - s(e - \epsilon))N) \rightarrow 0,$$

as  $N \rightarrow \infty$  for any  $\epsilon > 0$ . This shows that NMW holds for  $\beta$ -trumping channels in the micro-canonical case. Note that when we allow the catalyst to become correlated, NMW can be violated for microcanonical initial states, as shown above. This already makes clear that correlated catalysts provide a strict advantage in this set-up.

iii) Finally, let us also show that if we assume that  $\mathcal{N}$  is *unitary*, as we do in the case of  $\beta$ -catalytic channels, then NMW cannot be violated if the catalyst remains uncorrelated. The reason is the following: Since the global transformation on system and catalyst is unitary, it leaves the spectrum invariant. Since the catalyst remains invariant and uncorrelated, this implies that already the spectrum of the initial density matrix on the system has to remain invariant. Therefore there exists a unitary  $V$ , such that

$\mathcal{T}[\omega_\beta(H)] = V\omega_\beta(H)V^\dagger$ . As argued in case ii), we then have

$$\begin{aligned} p(w \geq \epsilon) &\approx \text{Tr}[P\mathcal{T}[\omega_\beta(H)]] = \text{Tr}[PV\omega_\beta(H)V^\dagger] \\ &\leq \text{Tr}[P\omega_\beta(H)], \end{aligned}$$

where the last inequality follows because Gibbs states are passive states and the first approximation holds to arbitrary accuracy as  $N \rightarrow \infty$ . However, by the same concentration inequalities we used to prove of our main results, we have

$$\text{Tr}[P\omega_\beta(H)] \leq K \exp(-k(\epsilon^2 N)^{1/(1+D)}),$$

for a non-critical many-body system in  $D$  spatial dimensions (with constants  $k, K > 0$ ). Thus, NMW holds true in this case as well.

## PARTIAL INFORMATION AND THE CANONICAL ENSEMBLE

---

### 6.1 THERMAL OPERATIONS UNDER PARTIAL INFORMATION

In this last chapter, we return to the theory of thermal operations. There are two seemingly independent questions that might be asked about this framework, each interesting in its own right: 1. Why do we assume that the initial state of the heat bath is described by a Gibbs state? 2. What are the state transitions that one can realise with thermal operations if one only has partial knowledge about the actual state of both the system and the heat bath? To motivate the first question is simple: A framework is only as useful as its constituting assumptions are sound. Hence, if thermal operations are to provide an adequate theoretical model of the thermodynamic evolution of a system, then the Gibbs state better be an adequate state representation of a system in thermal equilibrium. Now, empirically, this assumption has been greatly successful, but it would be great to be able to provide a *constructive* explanation of this success, rather than only having the empirical success justify the assumption *ex post*.

The second question, in turn, can easily be motivated as reflecting a much more realistic and common empirical scenario in thermodynamics than the original framework of thermal operations: Knowing the full “microstate”  $\rho$  of a system is often practically impossible or at least prohibitively expensive. Moreover, thermal operations, in allowing for any energy-preserving unitary on system and environment, assume an unrealistically high degree of control over the system and its environment. More often one may know only know the average energy of a state — based for example on prior energy measurements — and no further details about the underlying microstate. In such situations, thermal operations are not useful. The latter have been devised to derive fundamental bounds on the possible thermodynamical evolution of systems under the assumption that full control and knowledge is had about system and bath. In contrast, in the above situation, a much more helpful framework would be ones in which the achievable state transitions reflect the partial information about the system that one has and also the limited control that were originally responsible for only having partial information about the system. This is exactly the framework that we develop in the publication that forms the bulk of this chapter [4]. As such, the work presented here naturally fits into a line of research that attempts to bring the framework of thermal operations closer to the realities of experiment [129, 130, 131].

### 6.2 ROADS TO THE CANONICAL ENSEMBLE

Now, as it turns out, the answer to the second question also provides an interesting answer to the first question! Indeed, it provides a natural answer to a much more general question, namely why representing systems in statistical thermodynamics as canonical ensembles when their microstate is unknown is empirically as successful as it is. As such, the framework we develop here lets us contribute to an old question in the foundations of statistical thermodynamics. To better understand this contribution, it makes sense to briefly present some existing approaches to this question.

One common way to motivate the use of the canonical ensemble as the initial state of the heat bath in thermal operations is on the basis of *passivity*. An energy-incoherent state  $\rho$  is called passive with respect to some Hamiltonian  $H$  if higher-energy eigenstates carry less weight, that is, if the spectrum of  $\rho$  is ordered non-increasingly along higher energy. The thermal states can be singled out as the only family of states, whose members are *completely passive*, meaning that if  $\rho$  is passive with respect to  $H$ , then  $\rho^{\otimes n}$  is passive with respect to the Hamiltonian  $\sum_i^n H_i$  for all  $n \in \mathbb{N}$ , where  $H_i$  denotes the Hamiltonian that acts as  $H$  on the  $i$ th subsystem and trivially on all others. Operationally, passive states are interesting because no work can be extracted from them via a unitary process. The above implies that, if one used any other family of states to represent the initial state of the heat bath in thermal operations, then this would trivialize the state transitions  $\succeq_{TO}$ , because the “bath” could be used to extract work, in violation of the Second Law, and hence to prepare any final state [76]. While simple, the above argument is not very strong regarding the question of interest here, which is why the assignment of the canonical ensemble to thermodynamic systems is empirically successful. To begin with, it does nothing to motivate a physical mechanism for why the canonical ensemble is the proper state of the bath and why we don’t, empirically, find that we can in fact extract work from heat baths. It also does nothing to motivate why one should use *any* density operator to represent the initial state of the heat bath in the first place. Finally, it only applies to the question what the initial state of the *heat bath* should be, not how to represent the initial state of a system such as a working medium, that itself might not be in thermal equilibrium with the heat bath. This is because the argument relies on the assumption that one can extract work from a large number  $n$  of copies of the state  $\rho$ . But this is not the case, in general, where only one copy of  $\rho$  is had. Since we are concerned with answering the question not just for heat baths but also for other systems, the passivity argument does not make a strong case.

Another, more powerful approach to the question uses the notion of *typicality*. For instance, in canonical typicality [25, 26], the use of the canonical ensemble is essentially motivated by showing that, roughly speaking, if one was to sample uniformly from pure states in a microcanonical energy window of a large system whose spectrum is sufficiently widely “spread” over its subspaces, then the probability that one would sample a pure state that can locally be distinguished from a canonical ensemble is negligible. Hence, the argument goes, if one is uncertain about the underlying microstate of a system, then *if* one believes that the system dynamics are well described by a uniform measure over the pure state in the energy window, representing the state of the system as a Gibbs state is going to yield a good description with probability close to unity. This is a strong argument and it does not suffer from most of the problems from which the passivity argument suffers. However, the main drawback with it (and typicality arguments more generally) is that it is unclear just when the system dynamics are such that they are well described by a uniform measure over the pure states. Hence, in this argument the question is simply procrastinated to another level, namely the question why one state should be chosen over another is replaced by the question why one measure should be chosen over another when sampling from the microcanonical window (one way to argue for the uniform Haar measure over pure states is via the fundamental postulate of statistical mechanics, but this begs the question, since this postulate itself is something that is to be explained).

A third approach has first been articulated by Jaynes [158, 159] and goes by the name *Maximum entropy principle*. This principle states that, if one has only partial information about the state of a system, one should represent the system to be in that state with the largest Shannon (or, quantumly, von Neumann) entropy from the equivalence class of states whose

properties are compatible with the partial information. Jaynes argument for this principle is essentially methodological, rather than based on some physical mechanism. He argues that Shannon's axiomatic characterization of the Shannon entropy establishes this quantity as the unique, adequate measure of uncertainty, or partial knowledge. As such, the argument continues, the maximum entropy principle is the only rule to choose a state in light of uncertainty that is unbiased, the weakest commitment to a state in light of the available information. As such, the maximum entropy principle does not provide any physical explanation for why the rule should work out empirically in practice. Rather, Jaynes explains, whenever the empirical predictions based on the maximum entropy principle are empirically confirmed, then this shows that the partial information that was being held at the beginning completely captures the relevant dynamics at the level of empirical access. And this *latter* fact is related to the actual physical dynamics of the system of interest. As such, Jaynes' principle is a prescription that scientists should always follow, adapting the partial information that they use to decide a state representative until the predictions of the principle coincide with experimental observation. Jaynes' principle is often viewed critically as "subjectivist" and as relying too much on information-theoretic considerations and too little on the dynamics of physical systems. While some of these criticisms could be replied to along the lines of the argument above, it is certainly true that Jaynes' argument relies on the entropy as a special measure of uncertainty just like the typicality arguments rely on a particular choice of measure.

### 6.3 STATISTICAL ENSEMBLES WITHOUT TYPICALITY

In the following, we are going to present a framework in which we study the possible state transformations for a version of thermal operations in which agents only have partial information about the system and bath states. Our results provide a novel approach to the justification of the use of the canonical ensemble in statistical thermodynamics that arguably does without most of the drawbacks above. In particular, it does not rely on any special choice of measure or quantity. Instead, the result is *operational*, characterizing the canonical ensemble as that family of states that encodes the possible state transitions in thermal operations, if knowledge of all initial states is limited to their average energy.



ARTICLE

DOI: [10.1038/s41467-018-03230-y](https://doi.org/10.1038/s41467-018-03230-y)

OPEN

# Statistical ensembles without typicality

Paul Boes<sup>1</sup>, Henrik Wilming<sup>1</sup>, Jens Eisert<sup>1</sup> & Rodrigo Gallego<sup>1</sup>

Maximum-entropy ensembles are key primitives in statistical mechanics. Several approaches have been developed in order to justify the use of these ensembles in statistical descriptions. However, there is still no full consensus on the precise reasoning justifying the use of such ensembles. In this work, we provide an approach to derive maximum-entropy ensembles, taking a strictly operational perspective. We investigate the set of possible transitions that a system can undergo together with an environment, when one only has partial information about the system and its environment. The set of these transitions encodes thermodynamic laws and limitations on thermodynamic tasks as particular cases. Our main result is that the possible transitions are exactly those that are possible if both system and environment are assigned the maximum-entropy state compatible with the partial information. This justifies the overwhelming success of such ensembles and provides a derivation independent of typicality or information-theoretic measures.

---

<sup>1</sup>Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany. Correspondence and requests for materials should be addressed to P.B. (email: [pboes@zedat.fu-berlin.de](mailto:pboes@zedat.fu-berlin.de))

**M**aximum-entropy ensembles, such as the micro-canonical or the canonical ensemble, are the pillars on which statistical mechanics rests. Given some partial information about a system, a vast set of predictions about its behaviour can be derived by assigning to the system that statistical ensemble which maximises the entropy compatible with the partial information. Yet, in some ways this assignment may be seen as being peculiar in that there exist many other possible physical states that are compatible with this information. The assignment of maximum-entropy ensembles is primarily justified by its undoubtable empirical success when it comes to an agreement with experiment and observation. Thus, unsurprisingly, there has been much work aiming at providing theoretical grounds which explain its empirical success, going back to seminal work by Gibbs<sup>1</sup>. The most successful general arguments justifying the use of ensembles—both for classical and quantum systems—are either based on specific assumptions of the microscopic interactions from which ergodicity can be derived (see refs. 2,3 for a review on this approach and its conceptual problems), or based on the notion of typicality. The latter is the observation that the volume of pure quantum states (compatible with the information) that behave like a maximum-entropy ensemble is close to unity, with respect to a relevant measure on state space<sup>4–6</sup>. In these approaches, partially motivated by efforts in quantum thermodynamics<sup>7,8</sup>, the aim is to show that the system at hand behaves like the ensemble in the precise sense that it will output the same measurement statistics for a restricted, but most realistic and relevant, set of observables. In this way, the agreement between experiments and the assignment of ensembles is justified, with the only notorious problem that the measure that produces the typicality is difficult to justify dynamically. There have been attempts to derive precisely the emergence of canonical ensembles for most times from microscopic dynamical laws for common locally interacting quantum systems (for reviews, see refs. 9–11). However, it seems fair to say that it is still not fully clear yet why the probability of a system being, at any (or most) times, in a state should be described by this measure—a state of affairs particularly significant in light of the importance of this ensemble.

In this work, we provide a very different justification for the use of such ensembles. In contrast to the approaches mentioned before, our aim is not to derive that system's measurement statistics mimic those of the ensemble. Instead, we look at the possible state transitions that can be induced on a system from which one has only partial information (see also ref. 12). More precisely, we consider an initial system described only by partial information in the form of the expectation value of a set of observables. We pose the problem of finding the set of transitions that this initial system can undergo by evolving jointly with an environment when the state of this environment is itself known only partially, that is, up to expectation values with respect to a set of observables that correspond to those of the system. The environment plays the role of a usual heat bath and the set of transitions encode any possible task: extracting work, reaching a colder/warmer state, performing a computation or any other. Our main result is that, for any initial state, the possible state transitions on such a system under partial information coincide exactly with those possible if the system and the environment were initially in the maximum-entropy ensemble state compatible with the partial information. This then not only justifies the use of the canonical ensemble to represent a system under partial information, it also allows one to derive the building blocks of phenomenological thermodynamics without assuming systems to be represented by this ensemble. In fact our results can be seen as a derivation of the Gibbs entropy and the Clausius inequality without a priori assigning equilibrium states to the systems

involved. Finally, since our results hold for any initial state, they do not suffer from the problem of typicality approaches mentioned above and allow us to avoid assumptions about the system's Hilbert-space dimension (apart from being finite). In particular, our results also hold for small, individual quantum systems.

## Results

**Motivating example.** We begin the presentation of our setting with a motivating example. Consider a small quantum system  $S$  with Hamiltonian  $H$  within an environment  $E$  at temperature  $T$  and with Hamiltonian  $H_E$ , that is, an environment in the canonical ensemble at that temperature and Hamiltonian.

Given an initial quantum state  $\rho$  of the system, we can ask which final states of the system can be reached by coupling the system to the environment and evolving the joint system  $SE$  in such a way that the global entropy and energy remain unchanged, if one assumes perfect control over both the environment Hamiltonian  $H_E$  and the coupling, but for a fixed temperature  $T$ . Naturally, the answer to this question will strongly depend on the particular initial quantum state of  $S$ . For instance, the maximally mixed state  $\rho = \mathbb{1}_S$  and an energy eigenstate  $\rho' = |E_i\rangle\langle E_i|$  will generally allow for very different state transitions. That is, there will exist some final state  $\rho_f$  that can be reached by some entropy and energy preserving procedure  $O$  from  $\rho'$ , while no such procedure exists for  $\rho$ . Call this scenario the microstate scenario, because here one has full information about the actual 'microstates'—i.e. quantum states—of the system and the environment.

Suppose now that, instead of knowing the exact state of the system, one initially only knows its mean energy to be  $e$  with respect to  $H$ . We capture this partial information in what we call a macrostate of the form  $(e, H)$ . In this case, one can again ask which are the reachable states given that partial information. However, in this case the difficulty is that, in general, there will be many microstates compatible with this information. For instance, suppose that  $(e, H)$  is compatible with both  $\rho$  and  $\rho'$ . In this case  $\rho_f$  cannot be reached anymore because there is at least one state— $\rho$  in the previous example—compatible with the initial information for which  $\rho_f$  is unattainable. That said, one concludes that in order to reach some final state  $\rho_f$ , if only partial information about the initial state of  $S$  is had, one requires a single operational procedure  $O$  that takes any state compatible with the initial information to  $\rho_f$ . Note that this scenario is undesirably asymmetric in that the system's state is represented by a macrostate  $(e, H)$  (capturing our partial knowledge), while the environment microstate is fully known to be in the canonical ensemble at temperature  $T$ . Hence, one can go one step further and consider a situation in which not only does one only know the system's initial mean energy, but also the environment is described by a macrostate  $(e_E, H_E)$ . In this case, it becomes even more difficult to reach a given final microstate  $\rho_f$ , since now there has to exist a single procedure  $O$  that prepares  $\rho_f$  from any microstate of  $S$  compatible with  $e$  and any environment microstate compatible with  $e_E$ . Indeed, it may seem that in general no transition is possible under these circumstances. At the same time, this scenario most accurately describes the situation that one in fact faces in phenomenological thermodynamics, where only coarse-grained information is had about both system and environment. Call this last scenario then the macrostate scenario, because here both system and environment are described by macrostates  $(e, H)$  and  $(e_E, H_E)$ , respectively.

The main result of this work is to show that, not only do there exist possible transitions in the macrostate scenario, moreover these transitions are fully characterised by assigning maximum-



entropy ensembles to the macrostates involved: Under a natural model of operational procedures modelling thermodynamic transitions that we introduce below, given some value  $e$ , a final microstate  $\rho_f$  can be reached in the macrostate scenario if and only if it can be reached in the microstate scenario from the canonical ensemble state of energy  $e$ . Since the canonical ensemble is moreover the only state for which this equivalence holds, this result provides an explanation for the important role that the canonical ensemble plays in statistical mechanics, a theory formulated in the microstate scenario, to describe phenomenological thermodynamics, a theory formulated in the macrostate scenario.

**Formal setting.** We now proceed to make the notion of the microstate- and macrostate scenario rigorous and introduce our model of thermodynamic transitions, i.e. the transitions that a system  $S$  can undergo together with an arbitrary environment at fixed temperature.

Consider a  $d$ -dimensional quantum system  $S$  whose mean energy with respect to the Hamiltonian  $H$  is known to be  $e$ . We refer to the pair  $(e, H)$  as the 'macrostate' of the system, as it corresponds to a state of coarse-grained information about the system. Note, however, that we do not assume that the system is macroscopic, i.e. that  $d \gg 1$ . Every macrostate of the system corresponds to an equivalence class  $[e]_H$  of 'microstates'  $\rho \in D(\mathcal{H})$  of the system, namely all those density matrices whose mean energy with respect to  $H$  is  $e$ , with  $\mathcal{E}(\rho) := \text{tr}(\rho H) = e$ . The canonical ensemble corresponding to a macrostate  $(e, H)$  is then

$$\gamma_e(H) := \frac{e^{-\beta_S(e)H}}{\text{tr}(e^{-\beta_S(e)H})}, \tag{1}$$

where  $\beta_S(e)$  is chosen such that  $\text{tr}(\gamma_e(H)H) = e$ . Note that, by construction,  $\gamma_e$  is the maximum-entropy element in  $[e]_H$  and exists for every macrostate. As is clear from the example, in the following, we will often be concerned with making comparative statements about the microstate- and the macrostate scenarios. To simplify the presentation and highlight similarities between these scenarios, we now introduce the following convention: Let  $M$  be any map acting on microstates. Then  $M((e, H)) := M([e]_H)$  is the corresponding macrostate-level map. This notation will prove convenient in several ways. For instance, the requirement that an operation  $O$  maps all the states  $\rho$  compatible with  $(e, H)$  into the state  $\rho_f$  is simply expressed by

$$O((e, H)) = \rho_f. \tag{2}$$

Similarly, this notation can be also used to express operations on tensor products of macrostates. For instance, the expression

$$O((e, H) \otimes (e_E, H_E)) = \rho_f \tag{3}$$

implies that  $O(\rho \otimes \rho_E) = \rho_f$  for all  $\rho$  and  $\rho_E$  compatible with  $(e, H)$  and  $(e_E, H_E)$  respectively.

**Thermodynamic operations on macrostates.** Let us now describe and justify more precisely the form of a general macrostate operation as informally described above. With these operations we aim at capturing in full generality any possible transition that a system can undergo together with a heat bath. Hence, in order to describe an arbitrary macrostate operation, one is perfectly free to choose as an environment any system of arbitrary Hilbert-space dimension and with an arbitrary Hamiltonian  $H_E$ . As mentioned before, we do not assume that  $E$  is in a canonical ensemble—which would be fully determined by the inverse temperature  $\beta := (k_B T)^{-1}$ , dimension, and Hamiltonian—but to have a partial description in terms of its average energy,

thus assigning to it a macrostate  $(e_E, H_E)$ . We assume, as it is standard when considering thermodynamic operations<sup>13–15</sup>, that the system and the environment are initially uncorrelated, hence one initially possesses the macrostate compound  $(e, H) \otimes (e_E, H_E)$ . Naturally, the attachment of an uncorrelated environment can be iterated an arbitrary number of times, say  $N$ , bringing each time a new environment with an arbitrary dimension and Hamiltonian.

Moreover, since the macrostates provided by the environment model a bath, it is natural to assume that there exists a functional relationship between the environment Hamiltonian and the energy. In particular, we will assume this relationship to be that  $e_E = e_\beta(H_E)$ , where

$$e_\beta(H_E) := \text{tr}\left(\frac{e^{-\beta H_E}}{\text{tr}(e^{-\beta H_E})} H_E\right) \tag{4}$$

is the thermal energy of a bath at inverse temperature  $\beta$  and with Hamiltonian  $H_E$ . This assumption will be further discussed below. Dropping further the dependence on the Hamiltonian in (4) when it is clear from the context, the most general form of an initial macrostate then is of the form

$$(e, H) \otimes_{i=1}^N (e_\beta, H_{E_i}). \tag{5}$$

Given this model of the environment, we now turn to the describing the model of the joint evolution. Here, we aim at modelling the isolated evolution of SE, in the sense that it preserves the energy and entropy of the compound. Regarding the energy, one has to take into account that only mean values of the energy are accessible, hence it is most reasonable to impose only that the mean energy is preserved<sup>16–18</sup>, while noting that the mean energy must be preserved for all the initial microstates compatible with our initial macrostate (5). Regarding entropy conservation, we enforce it by imposing a unitary evolution of the compound. We note, however, that our results also hold for larger set of operations such as probabilistic mixtures of unitaries or entropy non-decreasing operations, or even more generally, any set of operations that contains unitary evolutions as a particular case.

Let us now, for sake of clarity, enumerate the assumptions that come into play when describing macrostate operations:

**Assumption 1: (Thermal energy environments)** Given an environment with Hamiltonian  $H_E$ , then the associated macrostate is given by  $(e_\beta(H_E), H_E)$ , where  $e_\beta(H_E)$  is the thermal energy at reference temperature  $T$ .

**Assumption 2: (Uncorrelated subsystems)** One can incorporate environmental systems that are initially uncorrelated with the initial system.

**Assumption 3: (Unitary evolution)** The compound SE undergoes a unitary evolution.

**Assumption 4: (Global mean energy conservation)** The unitary evolution of SE is such its mean energy is preserved for all the states (both of  $S$  and  $E$ ) compatible with our partial information.

Before turning to the formal definition of macrostate operations on the basis of these assumption, let us briefly comment on the assumption that environment macrostates have thermal energy (4). Clearly, this amounts to assume that environment macrostates have the same mean energy as the

canonical ensemble at inverse temperature  $\beta > 0$ ,

$$\gamma_\beta(H_E) := \frac{e^{-\beta H_E}}{\text{tr}(e^{-\beta H_E})}, \quad (6)$$

where we make the convenient abuse of notation of writing  $\beta$  directly as the subindex, unlike (1) where the mean energy was used instead. This is indeed unproblematic since  $e$  and  $\beta$  are in one to one correspondence, hence we will use  $\beta$  or  $e$  indistinctively when it is clear from the context.

We emphasise that (4) does not amount to assuming that the environment is in the canonical ensemble—which would beg the question by giving a prominent role to the canonical ensemble—since many states other than the canonical ensemble fulfilling (4) exist. Nevertheless, Assumption 1 could raise the criticism that our further results—the justification of ensembles—rely on a seemingly arbitrary energy assignment for the macrostate of E, as given by (4). However, we show in the Supplementary Methods 1 that (4) is the only possible assignment so that macrostate operations reflect indispensable features of thermodynamical operations. More precisely, we prove that (4) is the only energy function that does not allow one to extract an arbitrary amount of work from E alone—even if only partial information is given. Even more dramatically, it is the only energy function that does not trivialise macrostate operations, in the sense that any possible transition would be possible. Hence, (4) can be regarded as a necessary feature of an environment so that thermodynamic operations are sensibly accounted for in the formalism.

Finally, combining the notational convention for operations on macrostates, Assumptions 1–4, and denoting the global mean energy as  $\mathcal{E}(\rho_{SE}) := \text{tr}(\rho_{SE} H_{SE})$ , we define formally the set of macrostate operations with an environment at inverse temperature  $\beta$ :

**Definition 1: (Macrostate operations)** We say that  $\rho_f$  can be reached by macrostate operations from  $(e, H)$ , which we denote by

$$(e, H) \xrightarrow{\beta\text{-mac}} \rho_f, \quad (7)$$

if for any  $\epsilon > 0$  and  $\epsilon' > 0$  there exists an environment—that is, a set of  $N$  systems with respective Hamiltonians  $H_{E^i}$ —and a unitary

$U$  on SE, so that

$$\rho_f \approx_\epsilon \text{tr}_E \left( U(e, H) \bigotimes_{i=1}^N (e_{\beta, H_{E^i}}) U^\dagger \right) \quad (8)$$

while preserving the overall mean energy

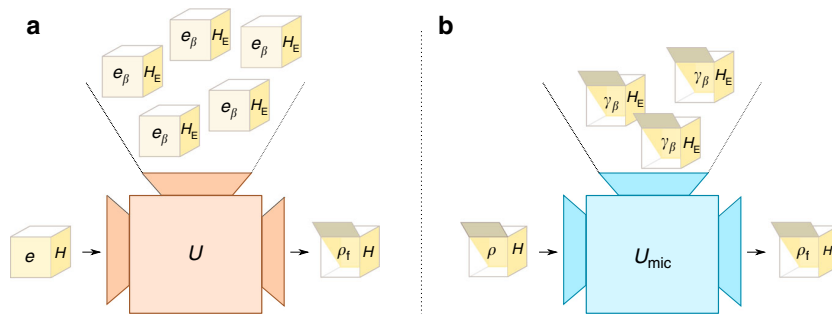
$$\mathcal{E} \left( U(e, H) \bigotimes_{i=1}^N (e_{\beta, H_{E^i}}) U^\dagger \right) \approx_{\epsilon'} \mathcal{E} \left( (e, H) \bigotimes_{i=1}^N (e_{\beta, H_{E^i}}) \right). \quad (9)$$

Here, we use  $\approx_\epsilon$  to say that two quantities differ by at most  $\epsilon$  in trace-norm, or in absolute value for expectation values. Note that although we allow for errors  $\epsilon, \epsilon'$  in the transition and in the mean energy conservation, those errors can be made arbitrarily small, hence it is for all practical purposes indistinguishable from an exact transition with exact mean energy conservation. It is also important to stress again that, in the previous definition and following the notation introduced with Eq. (3), both (8) and (9) have to be fulfilled for all the microstates compatible with the macrostates appearing in those equations. See Fig. 1a for a schematic description of macrostate operations as presented in Definition 1.

**Thermodynamic operations on microstates and main result.** As stated before, our main result consists in showing that not only is the set of reachable microstates under macrostate operations in general non-empty, it can also be characterised exactly by the corresponding canonical ensembles. In order to be able to state this correspondence between macrostates and their canonical ensembles formally, we will now introduce microstate operations as the corresponding model of thermodynamic transitions in the microstate scenario. These differ from macrostate operations only in that we assign a particular microstate to S and E. In other words, microstate operations are the complete analogue of the operations in Definition 1, but with full information about the actual quantum states involved. Hence, conditions (8) and (9) are modified, for microstate operations, in that they have to be fulfilled for a single state and not for a set of states compatible with our knowledge.

**Definition 2: (Microstate operations)** We say that  $\rho_f$  can be reached by microstate operations from  $\rho$ , which we denote by

$$\rho \xrightarrow{\beta\text{-mic}} \rho_f, \quad (10)$$



**Fig. 1** Pictorial representation of the equivalence between macrostate operations and microstate operations. Panel **a** shows macrostate operations and **b** microstate operation. Closed boxes represent systems from which we only know some partial information, in this case the mean energy. Inside the box there is the actual microstate unknown to us if the box is closed. Scenario **a** shows the situation where one has an initial system of which only the mean energy  $e$  is known and one can use any environment, being again limited to knowledge of its initial average energy  $e_\beta$ . The question is whether we can find a unitary  $U$  that takes the two systems, regardless of what is actually inside of them, to one box for which we are certain that we will find inside the microstate  $\rho_f$ . The answer to this question is provided by scenario **b**, where the initial boxes of system and environment are both open (implying that we know what is the microstate) and populated with the maximum-entropy ensemble.  $U$  exists if and only if there exists a unitary  $U_{\text{mic}}$  that implements the transition in **b** when taking  $\rho = \gamma_\beta(H)$ . This shows that a thermodynamic transition is possible if and only if it is also possible under the assignment of ensembles to systems

if for any  $\epsilon > 0$  and  $\epsilon' > 0$  there exists an environment—that is, a set of  $N$  systems with Hamiltonians  $H_{E^i}$ —and a unitary  $U$  on SE, so that

$$\rho_f \approx_\epsilon \text{tr}_E \left( U \rho \otimes_{i=1}^N \gamma_\beta(H_{E^i}) U^\dagger \right) \quad (11)$$

while preserving the overall mean energy

$$\mathcal{E} \left( U \rho \otimes_{i=1}^N \gamma_\beta(H_{E^i}) U^\dagger \right) \approx_{\epsilon'} \mathcal{E} \left( \rho \otimes_{i=1}^N \gamma_\beta(H_{E^i}) \right). \quad (12)$$

An operationally inspired illustration of the two types of operations as well as of our result is provided in Fig. 1.

In this setup, we call a macrostate  $(e, H)$  and a microstate  $\rho$  operationally equivalent, denoted as  $(e, H) \sim_\beta \rho$ , if

$$(e, H) \xrightarrow{\beta\text{-mac}} \rho_f \Leftrightarrow \rho \xrightarrow{\beta\text{-mic}} \rho_f. \quad (13)$$

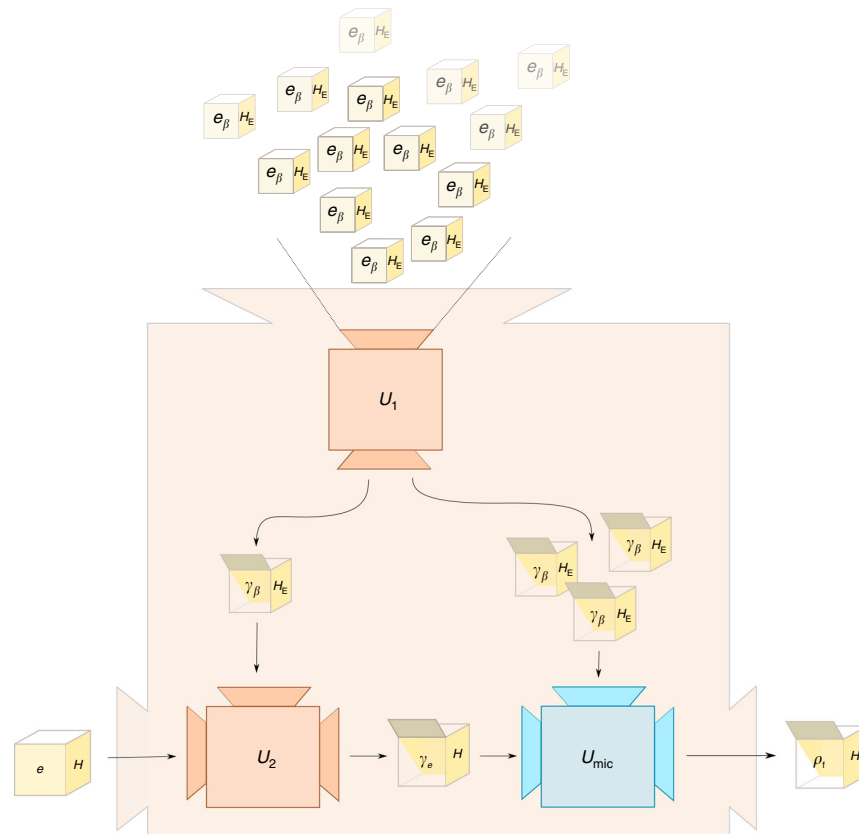
Whenever a macrostate and a microstate are related by the equivalence  $\sim_\beta$ , then, concerning the possible thermodynamic transitions, they are equivalent descriptions of the system. We are now in a position to state our main result.

**Theorem 3: (Equivalence with the canonical ensemble)** For any  $\beta \neq 0$ , the macrostate  $(e, H)$  is operationally equivalent to the corresponding canonical ensemble compatible with the partial information  $e$ . That is,

$$(e, H) \sim_\beta \gamma_e(H). \quad (14)$$

This theorem shows that, whenever the behaviour of a system under partial information concerns the possible thermodynamic transitions, a macrostate can be treated as if it was in its corresponding canonical ensemble, in the sense that their behaviours coincide exactly. A sketch of the proof, for illustration of the idea, is given in Fig. 2. The full proof appears in the Supplementary Methods 1.

Lastly, let us note that all of the above, including the operations and the notion of operational equivalence, can straightforwardly be generalised to the more general case of a set  $\mathcal{Q} = \{Q^j\}$  of  $n$  commuting observables replacing  $H$ , a vector  $\mathbf{v}$  of expectation values for each observable replacing  $e$  and by now parametrising the environment by a vector of inverse 'temperatures'  $\beta = (\beta^1, \dots, \beta^n)$  encoding other intensive quantities. In this case, we obtain an operational equivalence between the macrostate  $(\mathbf{v}, \mathcal{Q})$  and the corresponding maximum-entropy ensemble compatible with the partial information. More precisely, we obtain that, as long as



**Fig. 2** Sketch of proof of the main result. We show how an operation of the form of Fig. 1b can be used to build an operation of the form Fig. 1a. This gives the direction  $\Leftarrow$  in (13) for the equivalence of Theorem 3 (the other direction is trivial, see Supplementary Methods 1). The construction has three sub-blocks: Box  $U_1$  represents the fact that one can obtain the microstate  $\gamma_\beta(H_E)$  to arbitrary precision from many copies of the macrostate  $(e_\beta(H_E), H_E)$  using a macrostate operation (interestingly, this can be done with exact energy conservation). This result relies on a central limit theorem and typicality results for individual energy eigenspaces of many non-interacting systems. Box  $U_2$  operates by choosing as  $H_E$  as a rescaled version of  $H$  and showing that one can then obtain the microstate  $\gamma_e(H)$  using a macrostate operation. Box  $U_{mic}$  exists by assumption: it uses the microstate operation to obtain  $\rho_f$  from  $\gamma_e(H)$  (it is the one represented in Fig. 1b))

$\beta^j \neq 0$  for all  $j$ ,

$$(\mathbf{v}, \mathcal{Q}) \sim_{\beta} \gamma_{\mathbf{v}}(\mathcal{Q}), \quad (15)$$

where, in exact analogy to (1),  $\gamma_{\mathbf{v}}(\mathcal{Q})$  is the so-called generalised Gibbs ensemble (GGE)<sup>11,19–23</sup>

$$\gamma_{\mathbf{v}}(\mathcal{Q}) := \frac{e^{-\sum_j \beta^j_S(\mathbf{v}) Q^j}}{\text{tr} \left( e^{-\sum_j \beta^j_S(\mathbf{v}) Q^j} \right)}, \quad (16)$$

with  $\beta^j_S(\mathbf{v})$  being functions such that  $\text{tr}(Q^j \gamma_{\mathbf{v}}(\mathcal{Q})) = v^j$ . The scenario and derivation is completely analogous to that yielding Theorem 3 and it is presented in the Supplementary Methods 1.

**Rederiving bounds on work extraction.** At a conceptual level, we regard as our main contribution the theoretical justification, from an operational perspective, for the common and empirically extraordinarily well-supported use of the canonical ensembles in thermodynamics to describe systems in settings of partial information. The key step in this justification has been to prove a coincidence in behaviour with respect to thermodynamic transitions. The relevance of this coincidence is that many thermodynamic tasks and the laws of thermodynamics can ultimately be formulated as reflecting state transitions.

We illustrate this first using the task of work extraction and then derive the second law of thermodynamics in the form of the Clausius inequality.

Let us consider the following task: One is given a system S from which only the Hamiltonian and its mean energy  $e$  are given. For instance, S might be a burning fuel which one wants to use in a heat engine to perform work together with an environment. This common scenario is tackled in phenomenological thermodynamics by assigning to the system a temperature  $T_S$  and to the environment a temperature  $T$ . The optimal amount of work that can be performed is simply given by the difference of free energies of S during the process. Note that phenomenological thermodynamics operates at a level where only partial information—the thermodynamic variables—are given about both the system and the environment. Furthermore, the operation of such a heat engine is effectively independent of the precise microstate that describes S and E, exactly in the same spirit as that of Definition 1.

From the perspective of statistical mechanics, the assignment of a temperature  $T_S$  and  $T$  is understood as the assumption that both systems are in a canonical ensemble. Indeed, if we assume the system and the environment are initially in the state

$$\gamma_e \otimes \gamma_{\beta} := \gamma_e(H) \otimes \gamma_{\beta}(H_E) \quad (17)$$

one can formally derive limitations on the work  $\Delta W$ . The problem amounts to finding how much one can reduce the energy of the whole compound by any unitary operation that does not conserve the energy and assuming that all of the remaining energy can be extracted as work. One then obtains that this value is determined by the free energy as (see, e.g., ref. 16)

$$\begin{aligned} \Delta W^{\text{opt}} &:= \max_{U, H_E} \left[ \mathcal{E}(\gamma_e \otimes \gamma_{\beta}) - \mathcal{E}(U \gamma_e \otimes \gamma_{\beta} U^{\dagger}) \right] \\ &= \Delta \mathcal{E}_S - T \Delta \mathcal{S}_S := \Delta \mathcal{F}_S, \end{aligned} \quad (18)$$

where we denote the energy by  $\mathcal{E}(\rho_{SE}) = \text{tr}(\rho_{SE}(H_S + H_E))$ ,  $\Delta \mathcal{E}_S$  is the energy difference on S and  $\Delta \mathcal{S}_S$  is the difference of the von Neumann entropy on S. This yields the bound in terms of the free energy  $\mathcal{F}_S = \mathcal{E}_S - \beta^{-1} \mathcal{S}_S$  of the system and it relies only on the

first law of thermodynamics  $\Delta \mathcal{E}_{SE} = -\Delta W$  and the prescription of canonical ensembles to the system and environment.

We will now show that one can use Theorem 3 to derive the bound (18) without relying on the assumption (17) which assigns maximum-entropy ensembles to the systems at hand. The system S, given the partial information, is described by the macrostate  $(e, H)$ . We also have at our disposal an environment in any macrostate of the form  $\otimes (e_{\beta}(H_{E_i}), H_{E_i})$ . The goal is to perform work with a protocol in such a way that it achieves this work extraction for all possible microstates in the respective equivalence classes,  $[e]_H$  and  $[e_{\beta}(H_{E_i})]_{H_{E_i}}$  for all  $i$ , in a similar way to the way the laws of phenomenological thermodynamics allow one to extract work regardless of the actual microstates of the systems involved. It is clear that

$$\gamma_e(H) \xrightarrow{\beta\text{-mic}} \gamma_e(H) \forall e, H. \quad (19)$$

Hence, by invoking Theorem 3 one has also that

$$\begin{aligned} (e, H) &\xrightarrow{\beta\text{-mac}} \gamma_e(H), \\ (e_{\beta}(H_E), H_E) &\xrightarrow{\beta\text{-mac}} \gamma_{\beta}(H). \end{aligned} \quad (20)$$

Once we have the system S and the environment E in the states of at the r.h.s. of (20), we simply apply the unitary achieving the maximum in Eq. (18). In this way an amount of work given by  $\Delta \mathcal{F}_S$  is extracted. The fact that this is the optimal possible value that works for all microstates in  $[e]_H$  is trivial, since the work extraction has to be successfully implemented if the system is given in the state  $\gamma_e(H) \in [e]_H$ , for which the optimal value is  $\Delta \mathcal{F}_S$  as given by Eq. (18).

We conclude then that the optimal work that can be extracted from a system and an environment, from which we only know their mean energy coincides precisely with the optimal work when system and environment are described by their corresponding canonical ensemble. A completely analogous argument applies to any other conceivable task that can be formulated as concerning state transitions between microstates, both thermodynamically but also, and more generally, tasks with other conserved quantities.

**Second law and Clausius inequality.** Now we show that the second law of thermodynamics can be recovered by using Theorem 3. More particularly, we show that the set of achievable states  $\rho_f$  that can be reached by a transition of the form

$$(e, H) \xrightarrow{\beta\text{-mac}} \rho_f \quad (21)$$

can be determined only by merely taking into account the free energy  $\mathcal{F}$ . First note that by Theorem 3 the set of achievable  $\rho_f$  coincides with those that can be achieved by microstate operations of the form

$$\gamma_e(H) \xrightarrow{\beta\text{-mic}} \rho_f. \quad (22)$$

The set of achievable states by microstate operations has been investigated in ref. 16, where it is shown that the transition is possible if and only if the free energy decreases. Hence, we arrive at the second law of the form

$$(e, H) \xrightarrow{\beta\text{-mac}} \rho_f \Leftrightarrow \mathcal{F}(\gamma_e(H)) \geq \mathcal{F}(\rho_f). \quad (23)$$

Importantly, this result can also be seen as a derivation of the free energy as a state function  $F(e, H)$  on macrostates, by setting  $F(e, H) = \mathcal{F}(\gamma_e(H))$ . Since the energy is already naturally defined for macrostates we then also obtain the derived Gibbs entropy

$$S(e, H) := \frac{1}{T}(e - F(e, H)). \quad (24)$$

Interpreting the change of energy on the system as heat  $\Delta Q := e' - e$ , we see that a transition  $(e, H) \xrightarrow{\beta\text{-mac}} (e', H)$  between macrostates using macrostate operations is possible if and only if

$$\Delta Q \leq T\Delta S, \quad (25)$$

with  $\Delta S := S(e', H) - S(e, H)$ . We thus find that a state transition between macrostates is possible if and only if the Clausius inequality is fulfilled.

Lastly, we highlight that a generalisation of the same results for the case of multiple commuting observables is possible combining in a similar fashion Theorem 3 from the Supplementary Methods 1 with the results of ref. 17 to arrive at a formulation of the second law of the form

$$(\mathbf{v}, \mathcal{Q}) \xrightarrow{\beta\text{-mac}} \rho_f \Leftrightarrow \mathcal{G}(\gamma_{\mathbf{v}}(\mathcal{Q})) \geq \mathcal{G}(\rho_f), \quad (26)$$

where  $\mathcal{G}$  is the so-called free entropy defined as

$$\mathcal{G}(\rho) = \sum_j \beta_j \text{tr}(\rho Q^j) - S(\rho). \quad (27)$$

**Operational equivalence breaks for exact energy conservation.**

Theorem 3 establishes the operational equivalence between macrostates and their corresponding maximum-entropy ensembles based, among others, on Assumption 4, where it is assumed that the mean value of the energy is preserved. In this section, we consider a strengthening of macrostate and microstate operations in which Assumption 4 is replaced by the following:

Assumption 5: (Exact energy conservation) The unitary evolution  $U$  commutes with the total Hamiltonian,

$$[U, H_S + H_E] = 0. \quad (28)$$

We define, in full equivalence to the previous discussion, macrostate and microstate operations, but with exact preservation of the energy. We say that  $\rho_f$  can be reached by commuting macrostate operations from the macrostate  $(e, H)$ , similarly to Definition 1, but imposing, instead of mean energy conservation as in Eq. (9), the condition (28). One can define, analogously, commuting microstate operations by imposing similarly Eq. (28) and a notion of operational equivalence  $\overset{\sim}{\beta}$  analogous to (13).

In the Supplementary Methods 2, we show that for every  $\beta$  and non-trivial  $H$ , there exists at least one initial value  $e$ , such that

$$(e, H) \overset{\sim}{\beta} \gamma_e(H). \quad (29)$$

We believe the proof of this result to be interesting in its own right, because in it we show that the maps produced by commuting macrostate operations admit a simple linear characterisation, the details of which are discussed in the Supplementary Methods 2. We leave as a relevant open question to investigate particular cases where equivalence with the maximum-entropy ensemble is recovered for exact energy conservation. In the Supplementary Methods 3, we present one

setting in which the operational equivalence for the commuting case is recovered locally for large non-interacting systems. Another possibly fruitful direction is to impose extra restrictions on the set of possible states within an equivalence class and show equivalence only for this restricted class. Some partial results on this question are discussed in the Methods section. Also, note that (29) holds even if one replaces Assumption 1 for the assumption that the bath is already in a canonical ensemble. This follows since a bath fulfilling Assumption 1 can be transformed into a canonical state by unitaries that respect (28) (see Supplementary Methods 1). Again, an analogous breakdown of the equivalence as given by (29) exists for several commuting observables.

With respect to the justification of the use of maximum-entropy ensembles, this result implies that one cannot justify, in general, assigning a maximum-entropy state to a system under partial information by means of considering the possible thermodynamic transitions in a setting of exact energy conservation. This, we submit, again confirms current practice, because canonical ensembles are rarely used in situations where full control is had over the microdynamics of a system. From an operational point of view, note that Eq. (28) can be interpreted as system and bath being isolated from any other external system during the transition. However, in a situation where the only information available about the external system is also its mean energy, it seems challenging to certify that indeed system and bath evolved truly isolated. In this case, one can only be certain that the external system did not change its mean energy, which gives rise to the weaker condition of Assumption 4. Nonetheless, we regard both mean energy or exact energy conservation as reasonable assumptions whose adequacy will depend on the particular description of the situation at hand.

**The macroscopic limit.** In the light of the inequivalence of macrostates and their respective ensembles for the case of exact commutation, it is interesting to quantify by how much one has to violate (28) in order to recover equivalence. For this, let us introduce the random variable  $X$  which quantifies the energy change of SE during a macrostate operation. This energy change is captured by a probability distribution  $P$ . Theorem 3 implies the equivalence between the macrostate  $(e, H)$  and its corresponding ensemble with macrostate operations. These preserve the mean energy of the compound, hence with vanishing value of the first moment of  $P$ , although higher moments could well be different from zero. On the other hand, in the case of commuting macrostate operations, all the higher moments of  $P$  would indeed vanish due to condition (28). Hence, the deviation from zero of the higher moments of  $P$  seems a sensible quantifier of the violation of (28).

We will now discuss the behaviour of these higher moments for large, non-interacting and independent systems, capturing the classical limit of macroscopic systems. To do so, consider a system  $S$  described by  $N$  non-interacting subsystems. We will consider macrostate operations between a macrostate  $(e, H)$  and a final state  $\rho_f$  and impose that the final and initial states are large and uncorrelated. That is, instead of being any microstate in  $[e]_H$ , the initial microstate takes the form  $\sigma = \otimes_{i=1}^N \sigma^i$ . We also assume that the final state takes a similar form  $\rho_f = \otimes_i^N \rho_f^i$ . Using standard arguments of central limit theorems one can show that, in the limit of large  $N$  and for bounded Hamiltonians,  $P(X)$  for the transition  $(e, H) \xrightarrow{\text{mac}} \rho_f$  converges in distribution to a normal distribution with variance scaling as  $\sqrt{N}$ . Hence, the higher moments of  $P(X)$  per particle vanish (see Supplementary Methods 3). This is an argument in favour of the assignment of the ensemble to macrostates, for large weakly correlated systems, as long as one tolerates violations of (28)—as measured by the



higher moments—that are negligible in comparison with the typical energy scales involved in the thermodynamic operation.

**Comparison with existing work.** There exist several complementary approaches to justify the use of or single out maximum-entropy states in thermodynamics. As stated already in the introduction, the novelty of our approach lies in specifically assigning ensembles based on the set of possible thermodynamic transitions. This is in contrast with previous approaches, where canonical ensembles are justified based on measurement statistics of relevant observables. Both perspectives—the one presented here and previous approaches—can be fairly incorporated in a more general formulation about what is meant by a justification of the use of ensembles: the representation of a system's state by a statistical ensemble is justified with respect to some property if one can, on reasonable grounds, derive that the ensemble and the state behave exactly the same with respect to this property. Approaches based on notions of typicality usually consider as system states pure quantum states and the measurement statistics of some restricted set of observables—often local observables—as the property to be reproduced by the ensembles<sup>4,6</sup>. In contrast, in the present work, the system states are macrostates of partial information and the property is with respect to achievable state transitions under thermodynamic evolution. Theorem 3 justifies the assignment of maximum-entropy ensembles to macrostates with respect to such transitions. Macrostates are arguably the most common state assignment in thermodynamics, being at the root of discussions of the link of statistical mechanics and phenomenological thermodynamics, in that one often has knowledge of a system's state only up to its expectation values. Hence, this result provides a very broad operational justification of the use of maximum-entropy ensembles for a plethora of thermodynamical processes.

Another aspect that distinguishes our approach from other notions based on typicality is that we do not need to introduce a measure on quantum states or make any particular assumption on the dynamics. More precisely, known approaches based on typicality consider a given subset of quantum states and show that measurement statistics coincide with those of the ensemble for most of the quantum states within the subset. However, there is no general argument to advocate that one will find in nature precisely those states for which the statistics resemble those of the ensemble, even though these states comprise the vast majority according to reasonable measures. In contrast, one of the main features of our results is that it works for all and not for most of the quantum states that are compatible with the partial information. First, we demand that the transitions from macrostates, as given abstractly by (3), reach  $\rho_f$  for all the states compatible with the partial information. It would be analogous to the notion of typicality if we would instead demand that  $\rho_f$  is reached only from most of the microstates according to some state measure, but this is actually not required to derive our main results. Secondly, the equivalence between the macrostate and the corresponding ensemble holds for all possible macrostates, instead of just for a vast majority of the macrostate according to some measure on the possible values of the partial information. Most importantly, we stress that the equivalence between the macrostate and the ensemble holds irrespectively of the system's dimension. To put it in more practical terms, our results imply that a system, even if made of a few qubits, behaves as if it was in its maximum-entropy ensemble when it comes to state transitions under joint evolution with a possibly large bath. This is true in a single-shot regime—considering transitions on a single copy of the system at hand—without having to rely on taking the thermodynamic

limit where transitions of large number of copies are considered instead<sup>24,25</sup>.

Lastly, it may seem that our approach is closely related to that of the famous Jaynes' principle according to which a system should always be assigned the maximum-entropy state consistent with what one knows about it<sup>19,26</sup>. What both approaches have in common is that they consider the question of assigning microstates to macrostates. However, apart from this they differ considerably: Jaynes motivates his principle on the basis of Shannon's findings about the uniqueness of the Shannon entropy as an asymptotic measure of information. In contrast, our approach does not require us to assume any privileged measure of information, or even rely on any consideration about information measures at all. Moreover, as noted in the preceding paragraph, our approach also makes no reference to an asymptotic setting. Instead, in our work, we define a task on an individual system and investigate how an experimenter's partial knowledge about the system impacts her ability to execute this task. The canonical ensemble then naturally emerges as an effective representation of the experimenter's operational abilities in this setting. Again, no recourse to a measure of information, average performance, or even a subjectivist account of probabilities is required in our setting.

## Discussion

In this work, we have introduced a fresh way of justifying the very common use of maximum-entropy ensembles as a representation of the state of systems. We take a strictly operational stance to the subject, in which an experimenter has only partial information about the microstate of the system and all operations have to be compatible with such partial information. The vantage point for our argument concerns the possible thermodynamic transitions that systems can possibly undergo. This approach has the key advantages that it (a) naturally fits with many operational tasks in thermodynamics and its laws and (b) does not require underlying typicality arguments, and hence avoids some of their conceptual issues. We have also shown how our results can be used to derive features of phenomenological thermodynamics, such as the Gibbs entropy, free energy as state functions and the Clausius inequality, which determines whether a state transition on macrostates is possible without investing non-equilibrium resources. We are thus able to derive fundamental thermodynamic results without any assumption about typicality or information measures. Finally, our results generalise to the setting of several commuting observables. As such, the results here are likely to be of interest for thermodynamics in generalised settings or even outside the context of thermodynamics.

We point out as interesting further direction of research to incorporate probabilistic transitions to our formalism. We assume in our formalism that macroscopic operations transform any state of the equivalence class into a desired final state. It is an interesting endeavour to consider possible relaxations of this requirement by allowing some error probability on the transition. We leave it as an open question to investigate sets of reachable states under such relaxations. Lastly, the findings that operational equivalence breaks down for exact commutation suggest that further investigation is needed. In particular, it is natural to ask if one can impose additional constraints or assumptions to recover equivalence under exact energy conservation.

**Data availability.** Data sharing not applicable to this article as no data sets were generated or analysed during the current study.

Received: 6 October 2017 Accepted: 27 January 2018

Published online: 09 March 2018

**References**

- Gibbs, J. W. *Elementary Principles in Statistical Mechanics* (Chaeles Sribner's Sons, New York, 1902).
- Uffink, J. Compendium of the foundations of classical statistical physics in *Handbook for the Philosophy of Physics* (eds Earman, J. & Butterfield, J.) 924–1074 (North-Holland, Amsterdam, 2007).
- Haar, D. T. Foundations of statistical mechanics. *Rev. Mod. Phys.* **27**, 289–338 (1955).
- Goldstein, S., Lebowitz, J. L., Tumulka, R. & Zangh, N. Canonical typicality. *Phys. Rev. Lett.* **96**, 50403 (2006).
- Goldstein, S., Hara, T. & Tasaki, H. The second law of thermodynamics for pure quantum states. Preprint at <http://arXiv.org/abs/1303.6393> (2013).
- Popescu, S., Short, A. J. & Winter, A. Entanglement and the foundations of statistical mechanics. *Nat. Phys.* **2**, 754–758 (2006).
- Millen, J. & Xuereb, A. Perspective on quantum thermodynamics. *New. J. Phys.* **18**, 011002 (2016).
- Goold, J., Huber, M., Riera, A., del Rio, L. & Skrzypczyk, P. The role of quantum information in thermodynamics. *J. Phys. A* **49**, 143001 (2016).
- Eisert, J., Friesdorf, M. & Gogolin, C. Quantum many-body systems out of equilibrium. *Nat. Phys.* **11**, 124–130 (2015).
- Polkovnikov, A., Sengupta, K., Silva, A. & Vengalattore, M. Nonequilibrium dynamics of closed interacting quantum systems. *Rev. Mod. Phys.* **83**, 863–883 (2011).
- Gogolin, C. & Eisert, J. Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems. *Rep. Prog. Phys.* **79**, 56001 (2016).
- del Rio, L., Kraemer, L. & Renner, R. Resource theories of knowledge. Preprint at <http://arXiv.org/abs/1511.08818> (2015).
- Horodecki, M. & Oppenheim, J. Fundamental limitations for quantum and nanoscale thermodynamics. *Nat. Commun.* **4**, 2059 (2013).
- Brandão, F. G. S. L., Horodecki, M., Ng, N. H. Y., Oppenheim, J. & Wehner, S. The second laws of quantum thermodynamics. *Proc. Natl. Acad. Sci. USA* **112**, 3275–3279 (2015).
- Brandão, F. G. S. L., Horodecki, M., Oppenheim, J., Renes, J. M. & Spekkens, R. W. Resource theory of quantum states out of thermal equilibrium. *Phys. Rev. Lett.* **111**, 250404 (2013).
- Skrzypczyk, P., Short, A. J. & Popescu, S. Work extraction and thermodynamics for individual quantum systems. *Nat. Commun.* **5**, 4185 (2016).
- Guryanova, Y., Popescu, S., Short, A. J., Silva, R. & Skrzypczyk, P. Thermodynamics of quantum systems with multiple conserved quantities. *Nat. Commun.* **7**, 12049 (2016).
- Halpern, N. Y. Beyond heat baths II: Framework for generalized thermodynamic resource theories. *J. Phys. A* **51**, 094001 (2018).
- Jaynes, E. Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630 (1957).
- Rigol, M., Dunjko, V. & Olshanii, M. Thermalization and its mechanism for generic isolated quantum systems. *Nature* **452**, 854–858 (2008).
- Perarnau-Llobet, M., Riera, A., Gallego, R., Wilming, H. & Eisert, J. Work and entropy production in generalised Gibbs ensembles. *New. J. Phys.* **18**, 123035 (2016).
- Yunger Halpern, N., Faist, P., Oppenheim, J. & Winter, A. Microcanonical and resource-theoretic derivations of the thermal state of a quantum system with noncommuting charges. *Nat. Commun.* **7**, 12051 (2016).
- Lastaglio, M., Jennings, D. & Rudolph, T. Thermodynamic resource theories, non-commutativity and maximum entropy principles. *New. J. Phys.* **19**, 043008 (2017).
- Sparaciari, C., Oppenheim, J. & Fritz, T. Resource theory for work and heat. *Phys. Rev. A* **96**, 052112 (2017).
- Bera, M. N., Riera, A., Lewenstein, M. & Winter, A. Thermodynamics as a consequence of information conservation. Preprint at <http://arxiv.org/abs/1707.01750> (2017).
- Jaynes, E. Information theory and statistical mechanics. II. *Phys. Rev.* **108**, 171–190 (1957).

**Acknowledgements**

We thank H. Tasaki for comments. This work has been supported by the ERC (TAQ), the DFG (GA 2184/2-1, CRC 183, B02), the Studienstiftung des Deutschen Volkes, the EU (AQuS), and the COST action MP1209 on quantum thermodynamics.

**Author contributions**

P.B., H.W., J.E., and R.G. conceived the research question and wrote the article. P.B., H.W., and R.G. derived the technical results.


**Additional information**

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-03230-y>.

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://ngp.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

## **Supplementary Material - Statistical ensembles without typicality**

Boes et al.



**SUPPLEMENTARY METHODS 1**  
**PROOF OF OF MAIN RESULT**

**General maximum entropy ensembles**

In this section we generalize the formalism laid out in the results section of the main text to the case of many conserved quantities. That is, the macrostate and microstate operations and the notion of operational equivalence are generalised to the more general case of a set  $\{Q^j\}$  of  $n$  commuting observables replacing  $H$ , and a set  $\{v^j\}$  of expectation values for each observable replacing  $e$ . We introduce the following notation to arrange these sets into vectors

$$\mathcal{Q} = (Q^1, \dots, Q^n), \quad (1)$$

$$\mathbf{v} = (v^1, \dots, v^n), \quad (2)$$

so the macrostate of the system is given by  $(\mathbf{v}, \mathcal{Q})$ . The equivalence class of quantum states compatible with the macrostate is denoted by  $[\mathbf{v}]_{\mathcal{Q}}$ .

We model the environment with an analogous assumption as i) in the main text, but for the case of more conserved quantities. We assume that one can have access to  $N$  uncorrelated subsystems described each by a macrostate. The mean value of the conserved quantities is determined by the value of a vector of inverse ‘‘temperatures’’  $\beta = (\beta_1, \dots, \beta_n)$  for each conserved quantity. We denote, say for subsystem  $E^l$ , the conserved quantities and mean values as

$$\mathcal{Q}_{E^l} = (Q_{E^l}^1, \dots, Q_{E^l}^j), \quad (3)$$

$$\mathbf{v}_{\beta}(\mathcal{Q}_{E^l}) = (v_{\beta}(Q_{E^l}^1), \dots, v_{\beta}(Q_{E^l}^j)), \quad (4)$$

where we are making the slight abuse of notation to identify

$$Q_{E^l}^j := \mathbb{I}_1 \otimes \dots \otimes Q_{E^l}^j \otimes \dots \otimes \mathbb{I}_N. \quad (5)$$

In this way, we will denote the  $j$ -th conserved quantity on the whole environment as  $Q_E^j = \sum_{l=1}^N Q_{E^l}^j$ . Note that  $Q_E^j$  plays a similar role as the Hamiltonian of the environment  $H_E$  in the main text, but in this case for a different conserved quantity. Accordingly we also can arrange the conserved quantities of the environment, and the compound SE in a vector as

$$\mathcal{Q}_E = (Q_E^1, \dots, Q_E^n), \quad (6)$$

$$\mathcal{Q}_{SE} = (Q^1 + Q_E^1, \dots, Q^n + Q_E^n). \quad (7)$$

The environment is modeled by any macrostate of the form  $\bigotimes_{l=1}^N (\mathbf{v}_{\beta}(\mathcal{Q}_{E^l}), \mathcal{Q}_{E^l})$  where, in analogy to equation (4) of the main text, we assign a mean value of the conserved quantities equal to the ‘‘thermal’’ value, which in this case corresponds to the value that a maximum-entropy ensemble takes. That is,

$$v_{\beta}(Q_{E^l}^j) = \text{tr} \left( \gamma_{\beta}(\mathcal{Q}_{E^l}) Q_{E^l}^j \right), \quad (8)$$

where  $\gamma_{\beta}$  is the so-called generalised Gibbs ensemble (GGE) defined as

$$\gamma_{\beta}(\mathcal{Q}_{E^l}) := \frac{e^{-\sum_j \beta_j Q_{E^l}^j}}{\text{tr} \left( e^{-\sum_j \beta_j Q_{E^l}^j} \right)}. \quad (9)$$

We are now in a position to introduce macrostate operations.

*Definition 1* (Macrostate operations with many charges) We say that  $\rho_f$  can be reached by macrostate operations from  $(\mathbf{v}, \mathcal{Q})$ , which we denote by

$$(\mathbf{v}, \mathcal{Q}) \xrightarrow{\beta\text{-mac}} \rho_f, \quad (10)$$

if for any  $\epsilon > 0$  and  $\epsilon' > 0$  there exist an environment with observables  $\mathcal{Q}_E$ , and a unitary on SE such that

$$\left\| \text{tr}_E (U(\rho_i \otimes \rho_{E^1} \otimes \dots \otimes \rho_{E^m}) U^\dagger) - \rho_f \right\|_1 \leq \epsilon, \quad (11)$$

while preserving the global value of all the charges

$$\left| \text{tr} \left( U \left( \rho_i \bigotimes_{l=1}^N \rho_{E^l} \right) U^\dagger Q_{SE}^j \right) - \text{tr} \left( \rho_i \bigotimes_{l=1}^N \rho_{E^l} Q_{SE}^j \right) \right| \leq \epsilon', \quad (12)$$

for all  $j = 1, \dots, n$ . Importantly, both Supplementary Equations (11) and (12) have to be fulfilled for all the states of S and E compatible with our partial information, that is,

$$\forall \rho_i \in [\mathbf{v}]_{\mathcal{Q}}, \rho_{E^l} \in [\mathbf{v}_{\beta}(\mathcal{Q}_{E^l})]_{\mathcal{Q}_{E^l}} \text{ for } l \in [1, \dots, N]. \quad (13)$$

At this point, it is worth briefly discussing the physical significance of  $\mathcal{Q}$  and  $\mathcal{Q}_E$ . Our framework and in particular our main result – i.e. the equivalence with the maximum entropy ensemble presented in Theorem 3 of the main text – apply for any choice of charges for S and the environment E, given by  $\mathcal{Q}$  and  $\mathcal{Q}_E$  respectively, as long as the total mean value of the compound is preserved. In this sense our results leave open and completely general the choice of conserved quantities. However, one must be cautious by noting that imposing a conservation law of the mean value of  $\mathcal{Q} + \mathcal{Q}_E$  is not always well-justified. For instance, when  $\mathcal{Q}$  are the Hamiltonian, angular momentum and number of particles, it makes sense to allow for environments where  $\mathcal{Q}_{E^l}$  are the Hamiltonian, angular momentum and number of particles of  $E^l$  respectively. In this scenario, imposing Supplementary Equation (12) is meaningful. On the contrary if we take  $\mathcal{Q}$  to be the angular momentum and  $\mathcal{Q}_E$  to be, say, the magnetisation, we find that it might be in general unjustified to impose a conservation of  $\mathcal{Q} + \mathcal{Q}_E$ , since those two quantities are, a priori, unrelated. In summary, our framework takes as a starting point that a conservation law is imposed and builds upon this law. The prior arguments that justify imposing such a conservation law are outside the scope of this paper and must be considered independently.

The definition of  $\rho \xrightarrow{\beta\text{-mic}} \rho_f$  is completely analogous to the case of the previous section, with the GGE ensemble playing the role of the canonical ensemble.

*Definition 2* (Microstate operations with many charges) We say that  $\rho_f$  can be reached from  $\rho_i$  by microstate operations, which we denote by

$$\rho_i \xrightarrow{\beta\text{-mic}} \rho_f, \quad (14)$$

if for any  $\epsilon > 0$  and  $\epsilon' > 0$  there exist an environment with observables  $\mathcal{Q}_E$  and a unitary on SE such that

$$\| (U(\rho_i \otimes \gamma_{\beta}(\mathcal{Q}_E))U^{\dagger}) - \rho_f \|_1 \leq \epsilon, \quad (15)$$

while preserving the overall value of the charges

$$\left| \text{tr} \left( U(\rho_i \otimes \gamma_{\beta}(\mathcal{Q}_E))U^{\dagger} Q_{SE}^j \right) - \text{tr} \left( \rho_i \otimes \gamma_{\beta}(\mathcal{Q}_E) Q_{SE}^j \right) \right| \leq \epsilon', \quad (16)$$

for all  $j = 1, \dots, n$ .

We can now formulate the main result for the case of multiple observables:

*Theorem 3* (Equivalence with the GGE) Let  $\mathcal{Q}$  be any set of commuting observables and the environment be such that  $\beta^j \neq 0$  for all  $j$ . The macrostate  $(\mathbf{v}, \mathcal{Q})$  is operationally equivalent to the corresponding GGE ensemble compatible with the partial information  $\mathbf{v}$ . That is,

$$(\mathbf{v}, \mathcal{Q}) \sim_{\beta} \gamma_{\mathbf{v}}(\mathcal{Q}), \quad (17)$$

where  $\mathbf{v}$  are the inverse Lagrange multipliers that one assigns to S so that  $\text{tr}(Q^j \gamma_{\mathbf{v}}(\mathcal{Q})) = v^j$  for all  $j$ .

### Proof of Theorem 3

In this section, we will prove Theorem 3 above, which implies Theorem 3 in the main text as a special case. The equivalence relation in Supplementary Eq. (17) requires showing that

$$(\mathbf{v}, \mathcal{Q}) \xrightarrow{\beta\text{-mac}} \rho_f \Leftrightarrow \gamma_{\mathbf{v}}(\mathcal{Q}) \xrightarrow{\beta\text{-mic}} \rho_f. \quad (18)$$

The direction “ $\Rightarrow$ ” is trivial. Note that the l.h.s. implies that the transition is possible for all initial states compatible with  $(\mathbf{v}, \mathcal{Q})$ . In particular,  $\gamma_{\mathbf{v}}(\mathcal{Q})$  is one of these states compatible with  $(\mathbf{v}, \mathcal{Q})$  and hence the r.h.s. condition follows.

Before embarking on the proof of the direction “ $\Leftarrow$ ”, we will provide an overview of the different steps involved: 1. We show that macrostate operations allow us to consider without loss of generality probabilistic mixtures of unitary operations as well. 2. We show that using probabilistic mixtures of unitaries, we can reduce the problem to only considering microstates which are diagonal in the basis of the conserved quantities. 3. Using the previous results we show that we can “distill”, from the environment described by partial information, systems for which we are certain that they are in the microstates given by the GGE to arbitrary accuracy and with arbitrarily little change of the charges. This shows that we can effectively describe the

environment by GGE microstates directly. 4. We show that once we have an environment directly described by GGE microstates, we can always bring the system to the GGE microstate corresponding to its macrostate. That is, we show that it is possible to implement the transition

$$(\mathbf{v}, \mathcal{Q}) \xrightarrow{\beta\text{-mac}} \gamma_{\mathbf{v}}(\mathcal{Q}). \quad (19)$$

Finally, after we have replaced the state on the system with the GGE by a macrostate operation, we can apply the microstate operation that maps the GGE to the desired final state (r.h.s. of Supplementary Eq. (18) which is the premise of the proof). That is, we compose macrostate operations and microstate operations in the following way:

$$(\mathbf{v}, \mathcal{Q}) \xrightarrow{\beta\text{-mac}} \rho \wedge \rho \xrightarrow{\beta\text{-mic}} \sigma \Rightarrow (\mathbf{v}, \mathcal{Q}) \xrightarrow{\beta\text{-mac}} \sigma. \quad (20)$$

By taking  $\rho = \gamma_{\mathbf{v}}(\mathcal{Q})$  and  $\sigma = \rho_f$  and using Supplementary Eq. (19) we obtain the “ $\Leftarrow$ ”-direction, which concludes the proof. We will now give detailed derivations of steps 1.-4. separately.

### Reducing the problem to diagonal microstates

We now show that by being able to implement mixtures of energy-preserving unitaries, we can reduce the problem to one in which all microstates are diagonal in the eigenbasis of all the conserved quantities. To do that, define for every operator  $Q^j$  the mixture of unitaries

$$\rho \mapsto \mathcal{D}_{Q^j}(\rho) := \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T e^{iQ^j t} \rho e^{-iQ^j t} dt. \quad (21)$$

This mixture of unitaries dephases every state in the eigenbasis of  $Q^j$ . Since all the  $Q_j$  commute, we can sequentially apply these maps to map any state  $\rho \in [\mathbf{v}]_{\mathcal{Q}}$  to a state that commutes with all  $Q^j$ . In the following, we will denote this set of microstates that are diagonal in the eigenbasis of all the  $Q^j$  and correspond to the macrostate  $(\mathbf{v}, \mathcal{Q})$  by  $[\mathbf{v}]_{\mathcal{Q}}^{\text{diag}}$ . The fact that we can dephase all states without changing the mean values  $v^j$  implies that condition Supplementary Eq. (13) of Definition 1 can be relaxed to diagonal states, i.e.,

$$\forall \rho_l \in [\mathbf{v}]_{\mathcal{Q}}^{\text{diag}}, \rho_{E^l} \in [\mathbf{v}_{\beta}(\mathcal{Q}_{E^l})]_{\mathcal{Q}_{E^l}}^{\text{diag}} \text{ for } l \in [1, \dots, N]. \quad (22)$$

This allows us to restrict to diagonal states in the last two steps (3. and 4.).

### Mixtures of unitaries

We will now show that instead of considering unitary operations for macrostate operations, for finite temperature environments, we can also use probabilistic mixtures of unitaries. The basic idea is to use systems from the environment, described by the macrostate  $\bigotimes_{l=1}^N (\mathbf{v}_{\beta}(\mathcal{Q}_{E^l}), \mathcal{Q}_{E^l})$ , as a source of randomness.

Suppose we want to act with a mixture of unitaries on a system S at hand (which might include other systems from the environment). To do that, we first take two additional systems out of the environment. We choose these subsystems to be qubits labeled by  $E^1$  and  $E^2$  with  $\mathcal{Q}_{E^l} = (H, \mathbb{I}, \dots, \mathbb{I})$ . That is, we only consider the energy as a conserved quantity. Let us re-scale their Hamiltonian so that we can write it as  $H = 0|0\rangle\langle 0| + \Delta|1\rangle\langle 1|$ . As the macrostates have energy  $e_{\beta}(H)$  and they are uncorrelated this determines completely the diagonal of the microstates. One finds that if  $\rho_{E^1} \otimes \rho_{E^2} \in (e_{\beta}(H), H)^{\otimes 2}$  then  $\text{tr}(\rho_{E^1} \otimes \rho_{E^2} |i, j\rangle\langle i, j|) := p_{i,j} = p_i p_j$ , with  $p_1 = 1 - p_0 = e_{\beta}(H)/\Delta$ . Let us choose  $\Delta$  so that  $p_0 = 1/\sqrt{2}$ .

We now apply to the compound  $SE^1 E^2$  the unitary

$$U = |0, 0\rangle\langle 0, 0|_{E^1 E^2} \otimes U_S + (|0, 1\rangle\langle 0, 1| + |1, 0\rangle\langle 1, 0| + |1, 0\rangle\langle 1, 0|)_{E^1 E^2} \otimes U'_S. \quad (23)$$

One obtains that the effective map on the S is

$$\begin{aligned} \rho \mapsto \mathcal{M}(\rho) &= \text{tr}_{E^1 E^2}(U \rho U^\dagger) \\ &= p_{0,0} U_{\text{rest}} \rho U_{\text{rest}}^\dagger + (p_{0,1} + p_{1,0} + p_{1,1}) U'_{\text{rest}} \rho U_{\text{rest}}^\dagger \\ &= (p_0)^2 U_{\text{rest}} \rho U_{\text{rest}}^\dagger + (1 - (p_0)^2) U'_{\text{rest}} \rho U_{\text{rest}}^\dagger \\ &= \frac{1}{2} U_{\text{rest}} \rho U_{\text{rest}}^\dagger + \frac{1}{2} U'_{\text{rest}} \rho U_{\text{rest}}^\dagger. \end{aligned} \quad (24)$$

Repeating this process with as many pairs of qubits as required, we can apply any mixture of unitaries that we need. Hence, we can assume without loss of generality that in order to perform a macrostate operation as given by Definition 1, it suffices to find, instead of a single unitary  $U$  on the SE compound, a mixture of unitaries that performs the desired transition, which we denote as

$$\rho \mapsto \mathcal{U}(\rho) = \sum_{\lambda} p_{\lambda} U_{\lambda} \rho U_{\lambda}^{\dagger}, \quad (25)$$

with each of  $U_{\lambda}$  preserving the mean value of the conserved quantities.

### From the macrostate environment to the maximum entropy environment

The macrostate operations and the microstate operations employ different models of the environment. As discussed in the main text, the environment for macrostate operations is given by macrostates of the form

$$\bigotimes_{l=1}^N (\mathbf{v}_{\beta}(\mathcal{Q}_{E^l}), \mathcal{Q}_{E^l}). \quad (26)$$

On the other hand, for microstate operations one assumes that the environment is given by maximum entropy ensembles of the form

$$\bigotimes_{l=1}^{N'} \gamma_{\beta}(\mathcal{Q}_{E^l}). \quad (27)$$

We will now show that any environment of the form in Supplementary Eq. (27) can always be “distilled” from an environment of the form in Supplementary Eq. (26). That is, for any  $N'$  one can always find a sufficiently large  $N$  so that a system of the form in Supplementary Eq. (27) is obtained.

Due to the fact that we can implement mixtures of unitaries and dephase in the energy-eigenbasis we can, without loss of generality, model the macrostate operations that achieve this distillation by mixtures of unitaries that act on diagonal states of the bath, requiring only that they preserve the total expectation values of all the observables. For simplicity, we will take  $N' = 1$ , since an extension to larger values of  $N'$  can be done by simply repeating the process over  $N'$  copies of bath macrostates of the form in Supplementary Eq. (26).

For purely technical reasons, we will for now consider the special case where the eigenvalues of all the conserved quantities  $Q_{E^l}^j$  have rational eigenvalues. Since any operator can be approximated to arbitrary accuracy by one with rational eigenvalues, this is not a severe restriction.

Consider a larger number  $N$  of identical environment systems in the same macrostate  $(\mathbf{v}_{\beta}(\mathcal{Q}_{E^l}), \mathcal{Q}_{E^l}^l)$ , where  $\mathcal{Q}_{E^l} = \mathcal{Q}_{E^{l'}}$  for all  $l, l' = 1, \dots, N$ . We will apply a unitary map  $\mathcal{U}$  of the form in Supplementary Eq. (25) and find that the reduced state on every subsystem is given by  $\gamma_{\beta}(\mathcal{Q}_{E^l})$  to arbitrary accuracy as  $N \rightarrow \infty$ .

We first have to set up some notation. A basis-state on one of the subsystems can be labelled by the eigenvalues  $q_{\alpha}^j$  of the  $n$  conserved quantities  $Q_{E^l}^j$ , where  $\alpha = 1, \dots, d_{E^l}(j)$  and  $j = 1, \dots, n$ . Here,  $d_{E^l}(j)$  is the number of distinct eigenvalues of  $Q_{E^l}^j$ . Simplifying the notation, the basis states on system  $E^l$  can thus be labeled by  $d$  vectors  $\alpha^x = (\alpha_1^x, \dots, \alpha_n^x)$  corresponding to the choice of eigenvalues  $q_{\alpha_j^x}^j$ . A basis-state for the  $N$  systems is then given by choosing one vector  $\alpha^x$  for each subsystem and is denoted by  $\alpha^{\mathbf{x}} = (\alpha^{x_1}, \dots, \alpha^{x_N})$ . We will label the joint-eigenspaces of the  $Q_{E^l}^j$  on the  $N$  systems by  $\Pi_{\xi}$  and identify also  $\Pi_{\xi}$  with the projector onto that eigenspace. Given an eigenspace  $\Pi_{\xi}$ , we finally denote the corresponding eigenvalue of the total charge  $Q_{E^l}^j$  as  $q_{E,\xi}^j$ .

After setting up the notation, we will now start with the actual proof. The operation that we consider is very simple: We simply apply a completely random unitary in each of the subspaces  $\Pi_{\xi}$ . This operation clearly commutes with the total charges, hence it also preserves its average value. If we denote the total probability of subspace  $\Pi_{\xi}$  by  $p_{\xi}$ , it leaves the whole distribution  $p_{\xi}$  invariant, while leaving each of the subspaces in the maximally mixed state  $\Omega_{\xi}$ . Since each of the subspaces is permutation invariant, we find that the state of every system is finally described by the same density matrix

$$\rho'_{E^l} = \sum_{\xi} p_{\xi} \text{tr}_{\bar{l}}(\Omega_{\xi}). \quad (28)$$

Since the initial state  $\bigotimes_l \rho_{E^l}$  is uncorrelated, the total weight of joint eigenspaces  $\Pi_{\xi}$  for which any of the eigenvalues  $q_{E,\xi}^j$  deviates by more than  $O(\sqrt{N})$  from  $N v_{\beta}^j$  is exponentially small (by Hoeffding’s inequality). We will collect the remaining

subspaces in a set  $\mathcal{M}$ . We thus have

$$\rho'_{E^l} = \sum_{\xi \in \mathcal{M}} p_\xi \text{tr}_l(\Omega_\xi) + \epsilon_N \sigma, \quad (29)$$

where  $\sigma$  is some density-matrix and  $\epsilon_N$  goes to zero exponentially with  $N$ . Note that for all  $\xi \in \mathcal{M}$  the corresponding eigenvalues fulfill

$$|q_{E,\xi}^j/N - v_\beta^j| \leq \delta_N^j, \quad \delta_N^j \xrightarrow{N \rightarrow \infty} 0. \quad (30)$$

We will now show that, as  $N \rightarrow \infty$ , the reduced state on any single subsystem of each of the maximally mixed states  $\Omega_\xi$ , with  $\xi \in \mathcal{M}$ , approaches the GGE. To see this pick any such subspace  $\Pi_\xi$ . The fact that the eigenvalues  $q_\alpha^j$  are all rational, together with the fact that  $\xi \in \mathcal{M}$  implies that the dimension of any such subspace becomes arbitrarily large with increasing  $N$ .

Now consider the basis vectors  $\alpha^x = (\alpha^{x_1}, \dots, \alpha^{x_N})$  in  $\Pi_\xi$ . We will associate to each such basis vector a type

$$T(\alpha^x) = \left( \frac{k_1}{N}, \dots, \frac{k_d}{N} \right), \quad (31)$$

where  $k_x$  is the number of subsystems in state  $\alpha^x$ . In other words, they fulfill  $\sum_x k_x = N$  and

$$\sum_{x=1}^{d_{E^l}(j)} k_x q_{\alpha^x}^j = q_{E,\xi}^j. \quad (32)$$

The number of basis vectors corresponding to the same type  $T$  is given by

$$\#T = \frac{N!}{\prod_{x=1}^d k_x}. \quad (33)$$

It can be bounded using Stirling's approximation as

$$\sqrt{2\pi} \text{poly}(N) e^{NS(T)} \leq \#T \leq e \text{poly}(N) e^{NS(T)},$$

where  $S(T) = S(k_1/N, \dots, k_d/N)$  is the Shannon-entropy of a type. Note that the total dimension of one eigenspace  $\Pi_\xi$  is simply given by

$$d(\Pi_\xi) = \sum_{T \in \Pi_\xi} \#T. \quad (34)$$

A type has the property that  $T_x = k_x/N \geq 0$  and  $\sum_{x=1}^d k_x/N = 1$ . It can hence be interpreted as a probability distribution. If the total system is in the state  $\Omega_\xi$ , we obtain from permutation invariance that the probability to find the  $l$ -th subsystem in state  $\alpha^x$  is given by

$$p_{E^l}^\xi(\alpha^x) = \frac{\sum_{T \in \Pi_\xi} T_x \#T}{\sum_{T \in \Pi_\xi} \#T}. \quad (35)$$

We will now show that all types that differ from the GGE-distribution by more than  $\delta$  (in some norm on  $\mathbb{R}^{d-(n+1)}$ ) have a relative weight that vanishes as  $N \rightarrow \infty$ . In other words, as we increase the system size, the probability distribution  $p_{E^l}^\xi(\alpha^x)$  converges to that of a GGE with  $v^j = q_{E,\xi}^j/N$ . Let us denote the probability distribution corresponding to the GGE in subspace  $\xi$  by  $\gamma_\xi$ . Since the Shannon entropy is concave and has a unique maximum among all probability distributions compatible with the expectation values of the conserved quantities  $Q_{E^l}^j$  corresponding to the subspace  $\xi$ , we can bound the entropy of any type that differs by more than  $\delta$  from  $\gamma_\xi$  as

$$S(\gamma_\xi) - K'\delta^2 \leq S(T) \leq S(\gamma_\xi) - K\delta^2, \quad (36)$$

where the constants  $K$  and  $K'$  do not depend on  $N$ .

We thus see that the weight of the type is

$$\begin{aligned} \sqrt{2\pi} \text{poly}(N) e^{NS(\gamma_\xi) - NK'\delta^2} &\leq \#T \\ &\leq e \text{poly}(N) e^{NS(\gamma_\xi) - NK\delta^2}. \end{aligned}$$

Hence, the weight of the types is distributed according to a Gaussian-distribution on a subset of  $\mathbb{R}^{d-(n+1)}$  with variance  $\sigma^2$  of order  $1/N$ . For large  $N$ , it is thus very sharply peaked around the Gibbs-distribution and we can choose  $\delta$  to go to 0 as  $N \rightarrow \infty$  while at the same time most of the weight of the distribution is carried by distribution within  $\delta$  away from the GGE distribution. Choose, for example,  $\delta = N^{1/4}\sigma$ , so that

$$\lim_{N \rightarrow \infty} N^{1/4}\sigma = \lim_{N \rightarrow \infty} N^{1/4-1/2} = \lim_{n \rightarrow \infty} N^{-1/4} = 0. \quad (37)$$

More formally, we can upper bound the total weight of types more than  $\delta$  away from the GGE distribution by

$$\sum_{\substack{T \in \Pi_\xi, \\ \|T - \gamma_\xi\|_1 \geq \delta}} \#T \leq \mathcal{T}_\xi e \text{poly}(N) e^{NS(\gamma_\xi) - NK\delta^2}, \quad (38)$$

where  $\mathcal{T}_\xi$  is the total number of different types appearing in subspace  $\Pi_\xi$ . Similarly, for any  $q < 1$  we can lower bound the total weight of types closer than  $q\delta$  to the GGE distribution by

$$\sum_{\substack{T \in \Pi_\xi, \\ \|T - \gamma_\xi\|_1 \leq q\delta}} \#T \geq \text{poly}(q\delta) \mathcal{T}_\xi \sqrt{2\pi} \text{poly}(n) e^{NS(\gamma_\xi) - NK'q^2\delta^2}.$$

The relative volume of the two is then given by (using  $\delta = N^{-1/4}$ )

$$\begin{aligned} \frac{e \text{poly}(N) e^{NS(\gamma_\xi) - NK\delta^2}}{\sqrt{2\pi} \text{poly}(q\delta) \text{poly}(N) e^{NS(\gamma_\xi) - NK'q^2\delta^2}} &= \frac{e \text{poly}(N) e^{NS(\gamma_\xi) - \sqrt{N}K}}{\sqrt{2\pi} \text{poly}(qN^{-1/4}) \text{poly}(N) e^{NS(\gamma_\xi) - \sqrt{N}K'q^2}} \\ &\leq K'' \text{poly}(N) e^{-\sqrt{N}(K - K'q^2)} \rightarrow 0, \end{aligned} \quad (39)$$

for  $q < \sqrt{K/K'}$ . As  $N \rightarrow \infty$ , we therefore find that

$$\begin{aligned} \lim_{N \rightarrow \infty} \text{tr}_l(\Omega_\xi) &= \lim_{n \rightarrow \infty} \sum_x p_{E^l}^\xi(\alpha^x) |\alpha^x\rangle \langle \alpha^x| \\ &= \lim_{N \rightarrow \infty} \gamma_{\beta_\xi}(\mathcal{Q}_{E^l}) \\ &= \gamma_{\beta}(\mathcal{Q}_{E^l}), \end{aligned} \quad (40)$$

where  $\beta_\xi$  is the vector of "inverse temperatures" corresponding to the subspace  $\Pi_\xi$  and in the last line we have used that  $\lim_N q_{E,\xi}^j/N = \mathbf{v}_\beta^j$  for all  $\xi \in \mathcal{M}$ . Since this holds for all subspaces in  $\mathcal{M}$ , we finally obtain the desired result that

$$\rho'_{E^l} = \sum_{\xi \in \mathcal{M}} p_\xi \text{tr}_l(\Omega_\xi) + \epsilon_N \sigma \xrightarrow{N \rightarrow \infty} \gamma_{\beta}(\mathcal{Q}_{E^l}). \quad (41)$$

Concluding, we have shown that by taking many copies of the macrostate  $(\mathbf{v}_\beta, \mathcal{Q})$  and applying an exactly energy-conserving operation, we can prepare the microstate  $\gamma_{\beta}(\mathcal{Q})$ . Repeating this process many times, we can then also prepare any environment of the form

$$\bigotimes_l \gamma_{\beta}(\mathcal{Q}_{E^l}). \quad (42)$$

### Bringing the system to the maximum entropy state using the maximum entropy environment

In the last section we have proven that, from the model of the environment given by the form in Supplementary Eq. (26) for the definition of macrostate operations, one can distill a microstate environment of the form in Supplementary Eq. (27). We will now use such an environment to bring the system to the maximum entropy state. That is, to perform the transition in Supplementary Eq. (19). The idea to do that is very simple: We choose the right conserved quantities  $Q_E$  on the environment and then simply swap the system state with the environment.

Suppose that the system is in macrostate  $(\mathbf{v}, \mathcal{Q})$  with conserved quantities  $Q^j$  and let the corresponding inverse temperatures given by  $\gamma_{\mathbf{v}}(\mathcal{Q})$  be given by  $\beta_j(\mathbf{v})$ . Now choose the following conserved quantities on the environment,

$$Q_E^j = \frac{\beta_j(\mathbf{v})}{\beta_j} Q^j. \quad (43)$$

Of course, this is possible only if  $\beta_j \neq 0$  for all  $j$ . Then the two density matrices of the GGEs coincide,  $\gamma_\beta(Q_E) = \gamma_{\beta(v)}(Q)$ , and hence the total charge is conserved on average as the two states are swapped (it is not conserved exactly, since the microstate on the system can be any microstate in  $[v]_Q$ ). As mentioned in the previous section, the above reasoning strictly speaking only applies if the eigenvalues of  $Q_E^j$  are rational. However, we can always approximate  $Q_E^j$  by an operator with rational eigenvalues to arbitrary precision. In this case, the average charge conservation is fulfilled with arbitrary precision as well.

### Non-Gibbsian average energies trivialize thermodynamics

In this section, we will show that the assignment of macrostates to the environment as in Eq. (4) of the main text is the only one that does not lead to i) arbitrary work extraction from the environment and ii) trivial macrostate operations, in the sense that any transition is possible. For this, we will analyse the consequences of having an assignment of energies given by  $f(H)$  different from the one we assume for  $e_\beta(H)$ . For simplicity we will discuss it for the case of the energy as a single conserved quantity, since the argument is fully analogous for the case of other conserved quantities.

Let us first show i). The function  $f$  can, without loss of generality, be always expressed as  $f(H) = e_{\beta(H)}(H)$ , where now  $\beta(H)$  is not a fixed value but a function of the Hamiltonian. For the situation to not be equivalent to some fixed inverse temperature, at least two Hamiltonians must have different temperatures, i.e., there exist Hamiltonians  $H_1 \neq H_2$  such that  $\beta(H_1) \neq \beta(H_2)$ . For simplicity let us write  $\beta_j = \beta(H_j)$  in the following. Given any value of  $\beta_j$  we can distill, from a large number of macrostates of the environment, one canonical ensemble at temperature  $\beta_j$ . That is, from an environment of the form

$$\bigotimes_{j=1}^{N_1} (e_{\beta_1}(H_1), H_1) \bigotimes_{j=1}^{N_2} (e_{\beta_2}(H_2), H_2) \quad (44)$$

one can obtain systems in the microstate

$$\gamma_{\beta_1}(H_1)^{\otimes N'_1} \otimes \gamma_{\beta_2}(H_2)^{\otimes N'_2} \quad (45)$$

with  $N'_1$  and  $N'_2$  arbitrarily large for sufficiently large  $N_1$  and  $N_2$ . Once we possess two systems in the canonical ensemble at different inverse temperatures  $\beta_1$  and  $\beta_2$ , one can trivially extract work. That is, one could reduce the mean energy of Supplementary Eq. (45) and accumulate it in a work storage device. This is true since for some value for  $N'_1$  and  $N'_2$ , Supplementary Eq. (45) will cease to be a passive state [1].

The previous considerations imply trivially ii). Once we have established that the environment could be used to extract an arbitrary amount of work –mean energy–, one can invest this energy in creating an arbitrary state [2]. Hence one finds that if  $f(H)$  is not the thermal energy, then

$$(e, H) \xrightarrow{\beta\text{-mac}} \rho. \quad (46)$$

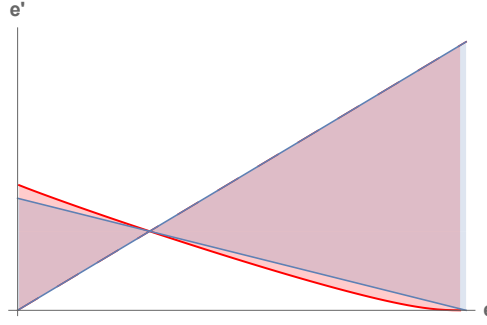
is possible for any  $\rho$ .

Altogether, we conclude that imposing that i) or ii) are impossible implies that  $f(H) = e_\beta(H)$  for a fixed  $\beta$ . In other words, there only exist specific families of functions, one for each value of  $\beta$ , that do not lead to trivial macrostate operations or work extraction from the environment. In this way the assignment of a parameter  $\beta$  to the environment follows from those basic principles. Importantly, note that the parameter  $\beta$  is in principle not related to any prior assignment of a temperature to the environment. For the sake of simplicity, we refer to  $\beta$  as the inverse temperature, but the interpretation of  $\beta$  as related to a prior value of  $T$  as  $\beta = (k_B T)^{-1}$  is not necessary to derive Theorem 3 of the Supplementary Material or any of the results in this work. In summary, we conclude that the only thermodynamically consistent way to assign average energies to environment systems is by assigning the energies corresponding to a thermal Gibbs state for some parameter  $\beta$  playing the role of an inverse temperature.

## SUPPLEMENTARY METHODS 2 BREAKDOWN OF EQUIVALENCE UNDER EXACT ENERGY CONSERVATION

In this section, we will prove the inequivalence between macrostates and their corresponding maximum-entropy ensemble when exact energy conservation, (28) in the main text, is imposed. In particular, we show that for every  $\beta$  and non-trivial  $H$ , there exists at least one initial value  $e$ , such that

$$(e, H) \not\approx_\beta \gamma_e(H). \quad (47)$$



**Figure 1:** Set of reachable final energies  $e'$ , given some Hamiltonian  $H$  and initial energy  $e$ . The reachable final energies under commuting macrostate operations are upper bounded by two lines (blue region) that themselves lower bound the set of reachable energies under microstate commuting operations (red region). The results of Ref. [4] imply that, for any non-trivial  $H$  and  $\beta$ , the red region has a non-linear boundary, which further implies that the blue region is strictly smaller than the red region. This, in turn, immediately gives Supplementary Eq. (52) and, hence, yields the breakdown of operational equivalence. In this figure, the intersection point marks the thermal energy  $e_{\beta}(H)$ , that is a fixed point of all operations by definition. Note further that the two sets are bounded, in one direction, by the identity. This follows from free energy considerations.

Let us first introduce some notation. We define commuting macrostate operations, denoted by

$$(e, H) \xrightarrow{\beta\text{-c-mac}} \rho_f, \quad (48)$$

similarly to Definition 1 of the main text but replacing condition (9) in the main text by  $[U, H_{SE}] = 0$ . In a similar fashion, we define commuting microstate operations, denoted by

$$\rho \xrightarrow{\beta\text{-c-mic}} \rho_f, \quad (49)$$

similarly to Definition 2 in the main text but replacing condition (12) in the main text by  $[U, H_{SE}] = 0$ . Commuting microstate operations are in the literature discussed as “thermal operations” [3, 4]. Proving the inequivalence in Supplementary Eq. (47) amounts to finding one microstate  $\sigma$  so that

$$(e, H) \xrightarrow{\beta\text{-c-mac}} \sigma, \quad (50)$$

$$\gamma_e(H) \xrightarrow{\beta\text{-c-mic}} \sigma. \quad (51)$$

The existence of such a state  $\sigma$  is implied by the fact that, for any non-trivial  $H \neq 0$  and any  $\beta$ , there exists at least one initial energy  $e$  such that

$$\max_{(e, H) \xrightarrow{\beta\text{-c-mac}} \rho_f} \mathcal{E}(\rho_f) < \max_{\gamma_e(H) \xrightarrow{\beta\text{-c-mic}} \rho_f} \mathcal{E}(\rho_f). \quad (52)$$

This equation implies the existence of  $\sigma$  because, if  $\sigma$  did not exist, then the reachable energies under the two types of operations would coincide. The equation itself follows from a result that we present in the next section and in which the reachable energies under macrostate commuting operations are linearly upper bounded, as illustrated in Fig. 1. We believe that this bound may be of independent interest.

### Partial characterisation of commuting macrostate transitions

In this section we will provide a method to analyse the allowed transitions under commuting macrostate operations. We cannot in general provide a full answer to which transitions  $(e, H) \xrightarrow{\beta\text{-c-mac}} \rho_f$  are possible. However, we will provide a method to bound the maximum and minimum energies of the states  $\rho_f$  achievable from a given macrostate  $(e, H)$ .

First, we need to consider a set of transitions between macrostates that are closely related to those produced by commuting macrostate operations:



*Definition 4* (Macrostate GP-maps) We say that  $(e', H)$  can be reached from  $(e, H)$  by macrostate GP-maps, which we denote by  $(e, H) \xrightarrow{\beta\text{-mGP}} (e', H)$ , if for any  $\epsilon > 0$  there exists a completely positive, trace preserving (CPTP)-map  $G$  such that

1.  $G(\gamma_\beta(H)) = \gamma_\beta(H)$ ,
2.  $G(\rho) \in [e']_\epsilon^H, \quad \forall \rho \in [e]_H$ .

Here,  $[e']_\epsilon^H$  denotes the union of the equivalence classes that differ from  $e'$  by at most  $\epsilon$ . By definition of the operations, and from results in Ref. [5], the following chain of implications holds: For any  $\rho \in [e']_H$ ,

$$(e, H) \xrightarrow{\beta\text{-c-mac}} \rho \Rightarrow (e, H) \xrightarrow{\beta\text{-mGP}} (e', H), \quad (53)$$

$$\Rightarrow \gamma_e(H) \xrightarrow{\beta\text{-c-mic}} \gamma_{e'}(H). \quad (54)$$

This in turn implies that for all  $(e, H)$ ,

$$\max_{(e, H) \xrightarrow{\beta\text{-c-mac}} \rho_f} \mathcal{E}(\rho_f) \leq \max_{(e, H) \xrightarrow{\beta\text{-mGP}} (e', H)} e' \leq \max_{\gamma_e(H) \xrightarrow{\beta\text{-c-mic}} \rho_f} \mathcal{E}(\rho_f). \quad (55)$$

From the results of Ref. [4] it follows that the rightmost term in the last equation is a non-linear function of  $e$ . In contrast, for the middle term, we find the following lemma.

*Lemma 5* (Reachable energies under macrostate GP-maps) For any non-trivial  $H$  and  $\beta$ , if  $e \in ]e_{\min}, e_{\max}[$ ,

$$\max_{(e, H) \xrightarrow{\beta\text{-mGP}} (e', H)} e' = \begin{cases} e & \text{if } e \geq e_\beta(H), \\ e_\beta(H) + \alpha(e)K_{\beta, H} & \text{if } e < e_\beta(H), \end{cases} \quad (56)$$

where  $e \mapsto \alpha(e)$  is a function linear in  $e$  and  $K_{\beta, H}$  is a constant independent of  $e$ . Similarly,

$$\min_{(e, H) \xrightarrow{\beta\text{-mGP}} (e', H)} e' = \begin{cases} e_\beta(H) + \alpha(e)K_{\beta, H} & \text{if } e \geq e_\beta(H), \\ e & \text{if } e < e_\beta(H). \end{cases} \quad (57)$$

This lemma characterizes the set of reachable energies under macrostate GP-maps, and hence upper and lower bounds the possible state transitions under commuting macrostate and microstate operations respectively. As discussed below, the constant  $K_{\beta, H}$  can easily be evaluated as a linear program. With respect to Supplementary Eq. (55), Lemma 5 and the results from Ref. [4] together imply that the second inequality in the equation has to be strict for at least one initial energy  $e \in ]e_{\min}, e_{\max}[$  and hence that Supplementary Eq. (52) holds.

### Proof of Lemma 5

Denote the set of macrostate GP-maps for a given initial energy  $e$  as  $\mathcal{G}_e$ . First, note that just like in the previous proofs, we need to consider only microstates  $\rho \in [e]_H^{\text{diag}}$  that are diagonal in the eigenbasis of  $H$ , because the decoherence map  $\mathcal{U}_{dec}$ . defined in Supplementary Eq. (21) is clearly a macrostate GP-map (mapping a macrostate to itself). Next, let

$$\mathcal{N} = \{A | \text{diag}(A) = A \wedge \text{tr}(H^\dagger A) = 0 \wedge \text{tr}(A) = 0\} \quad (58)$$

be the space of traceless, diagonal matrices that are orthogonal to  $H$ , for which  $\dim(\mathcal{N}) = d - 2$ . Further, let  $T$  be the matrix that is orthogonal to both  $H$  and  $\mathcal{N}$  and for which  $\text{tr}(H) = \text{tr}(T)$ . This matrix always exists. Clearly, if  $\{N_i\}_{i=1}^{d-2}$  is some orthogonal basis of  $\mathcal{N}$ , then  $\{H, T, N_1, \dots, N_{d-2}\}$  form a complete basis of the diagonal sector. For this reason, we can expand any diagonal state  $\rho$  as

$$\rho = \gamma_e(H) + \alpha(e)(H - T) + N(\rho), \quad (59)$$

where

$$\alpha(e) = \frac{e - e_\beta(H)}{\text{tr}(H^2)}, \quad (60)$$

$N(\rho) \in \mathcal{N}$ . Furthermore, by construction, in this expansion, any two states from the same equivalence class differ only by an element in  $\mathcal{N}$ . This expansion is useful because it allows us to show the following lemma.

**Lemma 6** (Characterising initial states in macrostate GP-maps) For non-trivial  $H$  and for any  $e \in ]e_{\min}, e_{\max}[$ , a CPTP-map satisfies condition 2 from Definition 4 iff  $G[\mathcal{N}] \subseteq \mathcal{N}$ .

*Proof.*  $\Leftarrow$ : Suppose there exists a map  $G$  and some state  $\rho \in [e]_H^{\text{diag}}$  such that

$$G(\rho) \in [e']_H. \quad (61)$$

If  $G[\mathcal{N}] \subseteq \mathcal{N}$ , then for any other state  $\rho' \in [e]_H^{\text{diag}}$ ,

$$\begin{aligned} \mathcal{E}(G(\rho')) &= \mathcal{E}(G(\rho)) + \mathcal{E}(G(N)) \\ &= e' + \mathcal{E}(N) \\ &= e', \end{aligned} \quad (62)$$

and hence  $G$  satisfies condition 2.

$\Rightarrow$ : Suppose that  $G \in \mathcal{G}_e$ . Then, for any  $\rho, \rho' \in [e]_H^{\text{diag}}$ , by Supplementary Eq. (59)

$$\rho - \rho' = N, \quad (63)$$

$$G(\rho) - G(\rho') = N', \quad (64)$$

and hence, by the linearity of CPTP-maps

$$\begin{aligned} G(N) &= G(\rho - \rho') \\ &= G(\rho) - G(\rho') \\ &= N'. \end{aligned} \quad (65)$$

This implies that  $G[\mathcal{N}_e] \subseteq \mathcal{N}$ , where

$$\mathcal{N}_e = \{N \in \mathcal{N} \mid \exists \rho, \rho' \in [e]_H^{\text{diag}} : \rho + N = \rho'\}. \quad (66)$$

$\mathcal{N}_e$  is the subspace of elements in  $\mathcal{N}$  that connect elements from  $[e]_H^{\text{diag}}$  with another. Now, for any  $e \in ]e_{\min}, e_{\max}[$ , that is, any non-extremal initial energy, it follows from the simplex geometry of the space of diagonal states that  $\dim(\mathcal{N}_e) = \dim([e]_H^{\text{diag}}) = \dim(\mathcal{N})$ . But this implies that there exists a complete  $((d-2)$ -dimensional basis)  $\{N_i\}$  of  $\mathcal{N}_e$  that also constitutes a basis for  $\mathcal{N}$ . Hence,  $G[\mathcal{N}_e] \subseteq \mathcal{N}$  implies  $G[\mathcal{N}] \subseteq \mathcal{N}$ .  $\square$

Note that for  $e \in \{e_{\min}, e_{\max}\}$ , depending on the degeneracy of the Hamiltonian  $H$ , it may be the case that  $\dim(\mathcal{N}_e) = \dim([e]_H^{\text{diag}}) = 0 \neq \dim(\mathcal{N})$ , so that the above lemma is not guaranteed to hold for extremal energies. Lastly, note that by the same reasoning, the proof holds also if we consider restricted sets of states strictly contained in  $[e]_H^{\text{diag}}$ , as long as the restricted set spans a vector space of the same dimensionality as the one spanned by  $[e]_H^{\text{diag}}$ . This has as a consequence that the breakdown of equivalence as phrased in the previous section as well as the results in this section hold if one further restricts the set of possible states in the equivalence class to an  $\epsilon$ -ball around a given state, which also spans a vector space of the right dimensionality.

To proceed, a corollary of Lemma 6 is that the set of macrostate GP-maps is the same, regardless of the initial energy, i.e.  $\mathcal{G}_e = \mathcal{G}_{e'}$ , for any  $e, e' \in ]e_{\min}, e_{\max}[$ . This allows us to drop the index in the following. Then, by Supplementary Eq. (59) we have

$$\begin{aligned} \max_{(e,H) \xrightarrow{\beta\text{-mGP}} (e',H)} e' &= \max_{G \in \mathcal{G}} \mathcal{E}(G(\rho)), \rho \in [e]_H^{\text{diag}} \\ &= \max_{G \in \mathcal{G}} \mathcal{E}(\gamma_e(H) + \alpha(e)G(H-T) + G(N(\rho))) \\ &= e_\beta(H) + \max_{G \in \mathcal{G}} \alpha(e)\mathcal{E}(G(H-T)). \end{aligned} \quad (67)$$

Finally, note that

$$\begin{aligned} &\max_{G \in \mathcal{G}} \alpha(e)\mathcal{E}(G(H-T)) = \\ &\begin{cases} \alpha(e) \max_{g \in \mathcal{G}} \mathcal{E}(G(H-T)), & \text{if } e \geq e_\beta(H), \\ \alpha(e) \min_{g \in \mathcal{G}} \mathcal{E}(G(H-T)), & \text{if } e < e_\beta(H), \end{cases} \end{aligned} \quad (68)$$

because  $\alpha(e)$  flips sign around  $e_\beta(H)$ . Defining the constants

$$F_{\beta,H} = \max_{g \in \mathcal{G}} \mathcal{E}(G(H - T)), \quad (69)$$

$$K_{\beta,H} = \min_{g \in \mathcal{G}} \mathcal{E}(G(H - T)), \quad (70)$$

we then have

$$\max_{(e,H) \xrightarrow{\beta\text{-mGP}} (e',H)} e' = \begin{cases} e_\beta(H) + \alpha(e)F_{\beta,H}, & \text{if } e \geq e_\beta(H), \\ e_\beta(H) + \alpha(e)K_{\beta,H}, & \text{if } e < e_\beta(H). \end{cases} \quad (71)$$

Similarly,

$$\min_{(e,H) \xrightarrow{\beta\text{-mGP}} (e',H)} e' = \begin{cases} e_\beta(H) + \alpha(e)K_{\beta,H}, & \text{if } e \geq e_\beta(H), \\ e_\beta(H) + \alpha(e)F_{\beta,H}, & \text{if } e < e_\beta(H). \end{cases} \quad (72)$$

In the final step, we will now discuss the values of  $F_{\beta,H}$  and  $K_{\beta,H}$ . The former can be found analytically to be such that

$$e_\beta(H) + \alpha(e)F_{\beta,H} = e. \quad (73)$$

To see this, note that the upper term in Supplementary Eq. (71) denotes the maximum reachable energy if the initial energy lies above the thermal energy (see Fig. 1). This is trivially is at least  $e$  (because the identity is always a macrostate GP-map). Now, if it was the case that

$$e_\beta(H) + \alpha(e)F_{\beta,H} > e, \quad (74)$$

then this would imply that there exists a GP-map  $G$  such that

$$\mathcal{E}(G(\gamma_e(H))) > e. \quad (75)$$

In this case,  $G$  would have certainly increased the free energy  $\Delta F(\rho) := S(\rho || \gamma_\beta(H))$  of the system, by monotonicity of the free energy of thermal states in  $e$ : For any  $e' > e, \rho \in [e']_H$ ,

$$\Delta F(\gamma_e(H)) < \Delta F(\gamma_{\beta_S(e')}(H)) \leq \Delta F(\rho). \quad (76)$$

Results from Ref. [5] imply that no GP-map can increase the free energy of the system, so that Supplementary Eq. (74) cannot be true, and hence  $F_{\beta,H}$  is determined by Supplementary Eq. (73).

Regarding  $K_{\beta,H}$ , it cannot in general be fixed analytically and depends on  $H$  and  $\beta$ . However, it can readily be computed with a linear program. This is because for any initial energy  $e$ , the optimization problems stated in Supplementary Equations (71) and (72) can be cast as linear programs. This is true since achievable state transitions under general GP-maps can be formulated as an LP [5, 6], and Lemma 6 shows that the only further constraint on macrostate GP-maps is itself linear, namely that  $\mathcal{G}(\mathcal{N}) \subseteq \mathcal{N}$ . Finally, note also that a similar Lemma to Lemma 6 can be shown to hold true for several commuting observables  $\mathcal{Q}$ . There, each of the observables  $Q^j$  is bounded linearly, so that, in total, the reachable states will be characterized by piece-wise linear bounds, instead of a single linear bound. Since this lemma is a straightforward generalization of Lemma 5, we omit its proof here.

### Local asymptotic equivalence in the commuting case

In this section we show that, locally and asymptotically, one can recover operational equivalence for the scenario in which both macrostate and microstate operations are commuting. Consider an  $N$ -partite, non-interacting system with initial macrostate  $\bigotimes_{l=1}^N (e, H)$ , that is, all parts share the same local Hamiltonian and initial energy. We will now use the results of the Supplementary Methods 1. There, it is shown that by means of commuting macrostate operation, one can bring a bath given by the macrostate of Supplementary Eq. (26) with  $N \rightarrow \infty$  to a final state  $\rho_f$  such that all its reduced states are arbitrarily close in trace norm to the maximum-entropy ensemble compatible with the partial information (see Supplementary Eq. (41)). We can now apply the same operation to the system in macrostate  $\bigotimes_{l=1}^N (e, H)$ . Formally,

$$\bigotimes_{l=1}^N (e, H) \xrightarrow{\beta\text{-c-mac}} \rho_f, \quad (77)$$

such that

$$\text{tr}_{\bar{l}}(\rho_f) \xrightarrow{N \rightarrow \infty} \gamma_e(H). \quad (78)$$

Note that the state  $\rho_f$  will be very different from the global canonical ensemble state  $\bigotimes_l \gamma_e(H)$  (that is reachable from the usual macrostate operations but not with commuting macrostate operations), since the different sites will in general be highly correlated.

Nevertheless, in direct analogy to the reasoning in the case of average-energy preserving operations, we can now apply any set of commuting microstate operations that act on the individual sites  $l$ , to prepare, locally, any state that could have been reached if instead one had started with the canonical ensemble states  $\gamma_e(H)$  on  $l$ . In this sense, local operational equivalence for the commuting case is recovered for i.i.d. and non-interacting systems in the thermodynamic limit. Note also that these results do not change if we additionally allowed access to Gibbs states on the bath, instead of macrostates, since we can again use the distillation procedure that we used before to arbitrarily well prepare Gibbs states  $\gamma$  using commuting operations. Also, as with the other results, this argument extends to the case of several commuting observables.

### SUPPLEMENTARY METHODS 3 MACROSTATE OPERATIONS IN THE MACROSCOPIC LIMIT

In this section we discuss the value of the higher moments of the energy difference  $X$  when performing a macrostate operation. As stated in the main text, we assume that  $H = \sum_i H^i$ . We first consider the case of a system whose subsystems are uncorrelated. That is, we assume the initial system macrostate to be of the form  $(e, H) = \bigotimes_{i=1}^N (e_i, H^i)$ . The canonical ensemble state for  $(e, H)$  is

$$\gamma_e(H) = \bigotimes_{i=1}^N \gamma_{\frac{e}{N}}(H^i). \quad (79)$$

Finally, we consider a macrostate transition  $(e, H) \xrightarrow{\beta\text{-mac}} \rho_f$ , where we also assume that

$$\rho_f = \bigotimes_{i=1}^N \rho_f^i. \quad (80)$$

We are interested in the distribution  $P(X)$ , where  $X$  is the change in energy under this macrostate transition.

To see that  $P$  will be normally distributed, we implement the above transition by acting on each of the subsystems independently. By Theorem 3 of the main text, we know that this is possible. In particular, by the procedure presented in the Supplementary Methods 1, we can implement the transition

$$(e_i, H^i) \xrightarrow{\beta\text{-mac}} \gamma_e(H^i) \quad (81)$$

as a macrostate transition, for any subsystem  $i$ . This produces a change in energy  $X_i$  with mean  $\mu_i$  and variance  $\sigma_i^2$ , which is finite for bounded  $H_i$ . Let  $s_N^2 = \sum_i \sigma_i^2$ . Then, by the Lyapunov Central Limit Theorem, we have that the total change in energy,  $X = \sum_i X_i$ , converges in distribution to a normal distribution,

$$\lim_{N \rightarrow \infty} X \xrightarrow{d} \mathcal{N}\left(\sum_i \mu_i = e' - e, s_N^2\right), \quad (82)$$

with  $e'$  being the final energy of the system, if the following condition is satisfied: There exists a  $\delta > 0$  such that

$$\lim_{N \rightarrow \infty} \frac{1}{s_N^{2+\delta}} \sum_i \mathbb{E}[|X_i - \mu_i|^{2+\delta}] = 0. \quad (83)$$

Choosing  $\delta = 1$  and since  $s_N^2 = O(N)$ , this is satisfied if  $\sum_i \mathbb{E}[|X_i - \mu_i|^{2+\delta}] = O(N)$ . This is a physically reasonable assumption to make. Now, from Supplementary Eq. (82) it follows that the energy change per subsystem is normally distributed as

$$\lim_{N \rightarrow \infty} \frac{X}{N} \xrightarrow{d} \mathcal{N}\left(e' - e, \frac{s_N^2}{N}\right). \quad (84)$$

In terms of the higher moments this means the following. Let

$$\mu_n(X) := \mathbb{E}[(X - \mu)^n], \quad n \in 1, 2, \dots \quad (85)$$

be the moments of a random variable  $X$ . If this  $X$  is normally distributed with variance  $\sigma^2$ , then independent of its mean the following is true and can be verified by evaluation.

$$\mu_{2n}(X) = \sigma^{2n}(2n - 1)!!, \quad \mu_{2n+1}(Y) = 0. \quad (86)$$

Combining this with Supplementary Eq. (84) we find that the higher moments per subsystem vanish in the macroscopic limit:

$$\lim_{N \rightarrow \infty} \mu_{2n}(X/N) = \lim_{N \rightarrow \infty} \left( \frac{s_N}{\sqrt{N}} \right)^{2n} (2n - 1)!! = 0. \quad (87)$$

As stated in the main text, this can be seen as an argument in favour of the assignment of the ensemble to macrostates, for large weakly-correlated systems, as long as one tolerates violations of (28) in the main text – as measured by the higher moments – that are negligible in comparison with the typical energy scales involved in the thermodynamic operation. Of course, a similar argument can be made for the case of weakly correlated systems. However, for conceptual clarity we here restricted to the independent case.

## REFERENCES

- [1] Pusz, W. & Woronowicz, S. L. Passive states and KMS states for general quantum systems. *Commun. Math. Phys.* **58**, 273–290 (1978).
- [2] Skrzypczyk, P., Short, A. J. & Popescu, S. Work extraction and thermodynamics for individual quantum systems. *Nature Comm.* **5**, 4185 (2016).
- [3] Brandão, F. G. S. L., Horodecki, M., Ng, N. H. Y., Oppenheim, J. & Wehner, S. The second laws of quantum thermodynamics. *PNAS* **112**, 3275–3279 (2015).
- [4] Horodecki, M. & Oppenheim, J. Fundamental limitations for quantum and nanoscale thermodynamics. *Nature Comm.* **4**, 2059 (2013).
- [5] Janzing, D., Wocjan, P., Zeier, R., Geiss, R. & Beth, T. Thermodynamic cost of reliability and low temperatures: Tightening Landauer’s principle and the second law. *Int. J. Th. Phys.* **39**, 2717–2753 (2000).
- [6] Renes, J. M. Work cost of thermal operations in quantum and nano thermodynamics. *Eur. J. Phys. Plus* **129**, 153 (2014).



## CONCLUSIONS

---

In this thesis, we have presented a systematic exposition of the resource theory of thermal operations that provides a formal framework to connect the postulates of quantum mechanics with the thermodynamics of both individual quantum systems, systems in the thermodynamic limit and at scales intermediate between these two extremes. We have then studied various modifications and generalizations of this framework to address a number of different questions of interest to the development of a theory of thermodynamics for small quantum systems. In particular, we have studied the extent to which access to randomness can be considered a resource for stochastic processes, establishing a strict separation between classical and quantum processes in terms of the required amount of randomness to realize a particular transition. We have also presented and studied in some depth the notion of correlated catalysis as a generalization of the more common notion of uncorrelated catalysis. This has been shown to be of considerable mathematical interest but also to allow for very interesting operational advantages in the context of work extraction and fluctuation theorems. Finally, we have used a model of thermal operations under partial information to single out the canonical ensemble as the unique microstate that encapsulates the possible state transitions that can be realized on systems and baths of which only their average energy is known (a statement that generalizes to other observables and ensembles).

These various findings all help develop the field of quantum thermodynamics, improve our understanding of the thermodynamics of small systems and help bring it from the days of its first conceptual breakthroughs to a mature theory that is not only of foundational interest but also ideally will lead, one day, to new quantum technologies. The works also open up several connections, albeit often not with the level of detail that the author would have wished, between quantum thermodynamics and related fields such as the theory of quantum computation, the classical theory of sampling complexity or to the statistical thermodynamics of out-of-equilibrium systems. It is our hope that the results presented here help others to make these connections clearer and to improve our understanding of the workings of the smallest machines that we can conceive.

### 7.1 OPEN QUESTIONS

#### *Catalytic quantum randomness*

The results presented in Chapter 4 derive tight bounds for the size of a source of randomness required to realize *any* transitions under majorization for a given system dimension. However, one can still ask for the required randomness of a *particular* transition  $\rho \rightarrow_{NO} \rho'$ . Preliminary investigations of this question suggest that this question is *NP*-hard, in the sense that the decision problem that takes an input  $(\rho, \rho', m)$  and decides whether there exists a noisy operation with a source of randomness of dimension  $m$  that realizes the above transition is *NP*-hard. This is indicated by the fact that the classical version of this problem can be reduced to the subset sum problem, which is *NP*-hard. However, it remains open to present a proper proof of this difficulty and to investigate the robustness of this hardness result.

A very different open question is to study whether one can use the key construction of the dephasing map presented in Chapter 4 to build interesting tensor networks, in particular a class of tensor networks known as *multi-scale entanglement renormalization ansatz (MERA)* [160, 161]. This class of tensor networks exhibits a characteristic causal structure in which isometries are used repeatedly to produce a particular entanglement structure across different subregions of a many-body system. Preliminary results suggest that using the unitary that implements the key construction in the above chapter as an isometry in a MERA network produces a very strong entanglement structure and could be used to define an analytically tractable network that saturates known bounds on the entanglement entropy of MERA networks. This remains to be investigated in more detail.

#### *Correlated Catalysis*

Here, there is one obvious open question, namely to decide the catalytic entropy conjecture that is presented in the corresponding chapter. To the best of the author's knowledge, the conjecture is open but interest is spreading through the community. A first step here would be to show that the set of states that are possible under trumping forms a subset of the set of states that are possible with a correlated catalyst. While this is almost certainly true, no proof of this statement is known to the author.

#### *Fluctuation theorems*

In the last part of Chapter 5.2, it was shown that correlated catalysts can be used in many-player strategies to engineer strongly correlated global work-distributions by having each of the players locally interact with the catalyst before passing it on. This is an intriguing feature of correlated catalysis that might potentially be of significant operational interest. However, a characterization of the kinds of distributions that can be engineered in this way is outstanding and seems to the author a promising research question.

#### *Canonical ensembles without typicality*

The findings of Chapter 6 might seem self-contained to the extent that no obvious follow-up questions to the project exist. However, to the author it seems that a key innovation of the paper is the notion of operational equivalence, in which two different resource theories are connected and compared by asking to what extent their elements operationally coincide. As such, it would be interesting to ask whether the notion of operational equivalence can be used in the context of other pairs of theories to make interesting connections, or possibly used to make statements at a more general level. For instance, one might try to provide an operational "derivation" of Jaynes' principle by asking: What are necessary and sufficient conditions on the properties of two resource theories such that (a) a unique family of states is picked out via operational equivalence for one of the theories and (b) that family is exactly the one that would be picked out by applying Jaynes' principle? In this context, it would be interesting to further investigate connection to a line of research in [51, 52], in which a "resource theory of knowledge" is developed and in which embeddings between resource theories at different levels of coarse-grainedness are studied.



## 7.2 ACKNOWLEDGEMENTS

My first thanks go to my supervisor Jens Eisert, who let me do this doctorate in his group and gave me a chance despite knowing that my previous focus on philosophy meant that I would have to spend the first years catching up. I always found the working environment in his large group – with many excellent researchers coming and going – extremely stimulating. I cannot overstate my thankfulness to Rodrigo Gallego and Henrik Wilming, both academically and personally. Whatever I might know about doing good science, I learned it from them (almost). I am thankful for their patience with me, especially for all the hours of their lives lost to me mumbling confused thoughts and scribbling ill-defined statements on the blackboard. I am very thankful to Nelly Ng, who was my go-to person for everything quantum thermo (and beyond) for the last years of my doctorate. Thanks to Henrik and Nelly for feedback on a draft of the thesis. Thanks to my additional co-authors Marcel Gohl, Markus Müller, Miguel Navascués, Swapan Rana, Carlo Sparaciari and Alexander Streltsov for working with me. Thanks also to the group members for all the great lunch conversations over the years, in particular Dominik Hangleiter, Christian Krumnow, Ingo Roth and Ryan Sweke. Thanks to Jörg Behrmann for keeping track of how much money I owe(d) him for coffee. I would also like to thank Felix von Oppen for agreeing to be the second examiner for this thesis, the Studienstiftung des deutschen Volkes for funding the first three years and the Templeton Foundation for funding the last year of this doctorate. Thanks to my family, especially my mother, for supporting me over the years of my studies. Finally, thanks to Babette for encouraging me whenever I learned that there was some mistake in my proof (again!). You were there for me whenever I needed you. This thesis is dedicated to Gebhard, my inspiring grandfather, and Marleen, my inspiring daughter. I wish they could have met.



## BIBLIOGRAPHY

- 
- [1] P. Boes, H. Wilming, R. Gallego, and J. Eisert, “Catalytic Quantum Randomness,” *Phys. Rev. X* **8**, 041016 (2018).
  - [2] P. Boes, J. Eisert, R. Gallego, M. P. Müller, and H. Wilming, “Von Neumann Entropy from Unitarity,” *Phys. Rev. Lett.* **122**, 210402 (2019).
  - [3] P. Boes, R. Gallego, N. H. Y. Ng, J. Eisert, and H. Wilming, “By-passing fluctuation theorems with a catalyst,” *Quantum* **4**, 231 (2020).
  - [4] P. Boes, H. Wilming, J. Eisert, and R. Gallego, “Statistical ensembles without typicality,” *Nat. Commun.* **9**, 1022 (2018).
  - [5] A. Streltsov, S. Rana, P. Boes, and J. Eisert, “Structure of the Resource Theory of Quantum Coherence,” *Phys. Rev. Lett.* **119**, 140402 (2017).
  - [6] P. Boës and M. Navascués, “Composing decoherence functionals,” *Phys. Rev. A* **95**, 022114 (2017).
  - [7] C. Sparaciari, M. Goihl, P. Boes, J. Eisert, and N. H. Y. Ng, “Bounding the resources for thermalizing many-body localized systems,” (2019), arXiv:1912.04920 [quant-ph] .
  - [8] P. Boes, H. Wilming, N. H. Y. Ng, and J. Eisert, “Second-order constraints for single-shot resource theories,” [in preparation] (2020).
  - [9] H. Wilming and P. Boes, “Separation of unital, exactly factorizable and random unitary channel via catalytic dilations,” [in preparation] (2020).
  - [10] D. A. Howard and M. Giovanelli, “Einstein’s philosophy of science,” in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta (Metaphysics Research Lab, Stanford University, 2019) fall 2019 ed.
  - [11] J. D. Bekenstein, “Black Holes and Entropy,” *Phys. Rev. D* **7**, 2333 (1973).
  - [12] A. R. Brown and L. Susskind, “Second law of quantum complexity,” *Phys. Rev. D* **97**, 086015 (2018).
  - [13] N. Perunov, R. A. Marsland, and J. L. England, “Statistical Physics of Adaptation,” *Phys. Rev. X* **6**, 021036 (2016).
  - [14] J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Chaeles Sribner’s Sons, New York, 1902).
  - [15] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal* **27**, 379 (1948).
  - [16] R. Landauer, “Irreversibility and heat generation in the computing process,” *IBM Journal of Research and Development* **5**, 183 (1961).
  - [17] C. H. Bennett, “Logical Reversibility of Computation,” *IBM J. Res. Dev.* **17**, 525 (1973).

- [18] C. H. Bennett, “*The thermodynamics of computation—a review*,” *Int. J. Theor. Phys.* **21**, 905 (1982).
- [19] J. Uffink, “Compendium of the foundations of classical statistical physics,” (2006), chapter for “*Handbook for Philosophy of Physics*”, J. Butterfield and J. Earman (eds).
- [20] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, “*Probability of second law violations in shearing steady states*,” *Phys. Rev. Lett.* **71**, 2401 (1993).
- [21] D. J. Evans and D. J. Searles, “*Equilibrium microstates which generate second law violating steady states*,” *Phys. Rev. E* **50**, 1645 (1994).
- [22] G. E. Crooks, “*Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences*,” *Phys. Rev. E* **60**, 2721 (1999).
- [23] C. Jarzynski, “*Nonequilibrium Equality for Free Energy Differences*,” *Phys. Rev. Lett.* **78**, 2690 (1997).
- [24] H. Tasaki, “Jarzynski relations for quantum systems and some applications,” (2000), arXiv:cond-mat/0009244 .
- [25] S. Goldstein, J. L. Lebowitz, R. Tumulka, and N. Zanghi, “*Canonical typicality*,” *Phys. Rev. Lett.* **96**, 50403 (2006).
- [26] S. Popescu, A. J. Short, and A. Winter, “*Entanglement and the foundations of statistical mechanics*,” *Nat. Phys.* **2**, 754 (2006).
- [27] C. Gogolin and J. Eisert, “*Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems*,” *Reports on Progress in Physics* **79**, 056001 (2016).
- [28] P. W. Anderson, “*Absence of diffusion in certain random lattices*,” *Phys. Rev.* **109**, 1492 (1958).
- [29] C. Gogolin, M. P. Müller, and J. Eisert, “*Absence of thermalization in nonintegrable systems*,” *Phys. Rev. Lett.* **106**, 40401 (2011).
- [30] W. Niedenzu, I. Mazets, G. Kurizki, and F. Jendrzejewski, “*Quantized refrigerator for an atomic cloud*,” *Quantum* **3**, 155 (2018).
- [31] K. Micadei, J. P. S. Peterson, A. M. Souza, R. S. Sarthour, I. S. Oliveira, G. T. Landi, T. B. Batalhão, R. M. Serra, and E. Lutz, “*Reversing the direction of heat flow using quantum correlations*,” *Nat. Commun.* **10**, 2456 (2019).
- [32] J. Klatzow, J. N. Becker, P. M. Ledingham, C. Weinzetl, K. T. Kaczmarek, D. J. Saunders, J. Nunn, I. A. Walmsley, R. Uzdin, and E. Poem, “*Experimental Demonstration of Quantum Effects in the Operation of Microscopic Heat Engines*,” *Phys. Rev. Lett.* **122**, 110601 (2019).
- [33] A. Ronzani, B. Karimi, J. Senior, Y.-C. Chang, J. T. Peltonen, C. Chen, and J. P. Pekola, “*Tunable photonic heat transport in a quantum heat valve*,” *Nat. Phys.* **14**, 991 (2018).
- [34] J. Goold, M. Huber, A. Riera, L. del Rio, and P. Skrzypczyk, “*The role of quantum information in thermodynamics—a topical review*,” *J. Phys. A Math. Theor.* **49**, 143001 (2016).
- [35] M. Lostaglio, “*An introductory review of the resource theory approach to thermodynamics*,” *Reports Prog. Phys.* **82**, 114001 (2019).

- [36] F. Binder, L. A. Correa, C. Gogolin, J. Anders, and G. Adesso, eds., *Thermodynamics in the Quantum Regime*, Fundamental Theories of Physics, Vol. 195 (Springer International Publishing, Cham, 2018).
- [37] S. Deffner and S. Campbell, *Quantum Thermodynamics*, 2053-2571 (Morgan and Claypool Publishers, 2019).
- [38] 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 (2015).
- [39] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).
- [40] A. S. L. Malabarba, L. P. García-Pintos, N. Linden, T. C. Farrelly, and A. J. Short, “Quantum systems equilibrate rapidly for most observables,” *Phys. Rev. E* **90**, 12121 (2014).
- [41] M. Horodecki and J. Oppenheim, “Fundamental limitations for quantum and nanoscale thermodynamics,” *Nat. Commun.* **4**, 2059 (2013).
- [42] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and T. Beth, “Thermodynamic Cost of Reliability and Low Temperatures: Tightening Landauer’s Principle and the Second Law,” *Int. J. Theor. Phys.* **39**, 2717 (2000).
- [43] 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,” *Phys. Rev. Lett.* **111**, 250404 (2013).
- [44] E. H. Lieb and J. Yngvason, “The physics and mathematics of the second law of thermodynamics,” *Phys. Rep.* **310**, 1 (1999).
- [45] E. H. Lieb and J. Yngvason, “The Mathematical Structure of the Second Law of Thermodynamics,” in *Current Developments in Mathematics, 2001* (International Press, Cambridge, 2002) pp. 89–130.
- [46] C. Carathéodory, “Untersuchungen über die Grundlagen der Thermodynamik,” *Math. Ann.* **67**, 355 (1909).
- [47] M. Weilenmann, L. Kraemer, P. Faist, and R. Renner, “Axiomatic relation between thermodynamic and information-theoretic entropies,” *Phys. Rev. Lett.* **117**, 260601 (2016).
- [48] F. G. S. L. Brandão and G. Gour, “Reversible framework for quantum resource theories,” *Phys. Rev. Lett.* **115**, 070503 (2015).
- [49] T. Fritz, “Resource convertibility and ordered commutative monoids,” (2015), arXiv:1504.03661 [math.OC] .
- [50] B. Coecke, T. Fritz, and R. W. Spekkens, “A mathematical theory of resources,” *Information and Computation* **250**, 59 (2016), quantum Physics and Logic.
- [51] L. del Rio, *Resource theories of knowledge*, Ph.D. thesis, ETH Zurich (2015).
- [52] L. Kraemer and L. del Rio, “Currencies in resource theories,” (2016), arXiv:1605.01064 [quant-ph] .
- [53] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, “Concentrating partial entanglement by local operations,” *Phys. Rev. A* **53**, 2046 (1996).

- [54] M. Horodecki and J. Oppenheim, “(Quantumness in the context of) resource theories,” *Int. J. Mod. Phys.* **27**, 1345019 (2012).
- [55] M. Horodecki, P. Horodecki, and J. Oppenheim, “Reversible transformations from pure to mixed states and the unique measure of information,” *Phys. Rev. A* **67**, 62104 (2003).
- [56] J. Åberg, “Quantifying superposition,” (2006), arXiv:quant-ph/0612146 [quant-ph].
- [57] T. Baumgratz, M. Cramer, and M. B. Plenio, “Quantifying coherence,” *Phys. Rev. Lett.* **113**, 140401 (2014).
- [58] I. Marvian and R. W. Spekkens, “Modes of asymmetry: The application of harmonic analysis to symmetric quantum dynamics and quantum reference frames,” *Phys. Rev. A* **90**, 062110 (2014).
- [59] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, “Reference frames, superselection rules, and quantum information,” *Rev. Mod. Phys.* **79**, 555 (2007).
- [60] V. Veitch, S. A. Hamed Mousavian, D. Gottesman, and J. Emerson, “The resource theory of stabilizer quantum computation,” *New J. Phys.* **16**, 013009 (2014).
- [61] E. Chitambar and G. Gour, “Quantum resource theories,” *Rev. Mod. Phys.* **91**, 025001 (2019).
- [62] N. Yunger Halpern and J. M. Renes, “Beyond heat baths: Generalized resource theories for small-scale thermodynamics,” *Phys. Rev. E* **93**, 022126 (2016).
- [63] N. Y. Halpern, “Beyond heat baths II: framework for generalized thermodynamic resource theories,” *Journal of Physics A: Mathematical and Theoretical* **51**, 094001 (2018).
- [64] N. Y. Halpern, P. Faist, J. Oppenheim, and A. Winter, “Microcanonical and resource-theoretic derivations of the thermal state of a quantum system with noncommuting charges,” *Nat. Commun.* **7**, 12051 (2016).
- [65] Y. Guryanova, S. Popescu, A. J. Short, R. Silva, and P. Skrzypczyk, “Thermodynamics of quantum systems with multiple conserved quantities,” *Nat. Commun.* **7**, 12049 (2016).
- [66] H. Wilming, *A Quantum of Thermodynamics*, Ph.D. thesis, Freie Universität Berlin (2018).
- [67] H. Joe, “Majorization and divergence,” *J. Math. Anal. Appl.* **148**, 287 (1990).
- [68] E. Ruch, R. Schranner, and T. H. Seligman, “The mixing distance,” *J. Chem. Phys.* **69**, 386 (1978).
- [69] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities : theory of majorization and its applications* (Springer Science+Business Media, LLC, 2011).
- [70] H. Wilming and R. Gallego, “Third Law of Thermodynamics as a Single Inequality,” *Phys. Rev. X* **7**, 041033 (2017).
- [71] I. Csiszár, “Axiomatic Characterizations of Information Measures,” *Entropy* **10**, 261 (2008).
- [72] L. Masanes and J. Oppenheim, “A general derivation and quantification of the third law of thermodynamics,” *Nat. Commun.* **8**, 14538 (2017).
- [73] J. Scharlau and M. P. Mueller, “Quantum Horn’s lemma, finite heat baths, and the third law of thermodynamics,” *Quantum* **2**, 54 (2018).

- [74] C. Sparaciari, L. del Rio, C. M. Scandolo, P. Faist, and J. Oppenheim, “The first law of general quantum resource theories,” (2018), arXiv:1806.04937 [quant-ph] .
- [75] P. Skrzypczyk, A. J. Short, and S. Popescu, “Work extraction and thermodynamics for individual quantum systems,” *Nat. Commun.* **5**, 2333 (2014).
- [76] 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,” *Phys. Rev. Lett.* **111**, 250404 (2013).
- [77] R. Gallego, J. Eisert, and H. Wilming, “Thermodynamic work from operational principles,” *New J. Phys.* **18**, 103017 (2016).
- [78] C. Sparaciari, J. Oppenheim, and T. Fritz, “Resource theory for work and heat,” *Phys. Rev. A* **96**, 052112 (2017).
- [79] C. T. Chubb, M. Tomamichel, and K. Korzekwa, “Beyond the thermodynamic limit: finite-size corrections to state interconversion rates,” *Quantum* **2**, 108 (2018).
- [80] A. Rényi, “On measures of entropy and information,” in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* (University of California Press, Berkeley, Calif., 1961) pp. 547–561.
- [81] M. Tomamichel, M. Berta, and J. M. Renes, “Quantum coding with finite resources,” *Nat. Commun.* **7**, 11419 (2016).
- [82] L. Wang and R. Renner, “One-shot classical-quantum capacity and hypothesis testing,” *Phys. Rev. Lett.* **108**, 200501 (2012).
- [83] F. Nielsen and S. Boltz, “The Burbea-Rao and Bhattacharyya Centroids,” *IEEE Trans. Inf. Theory* **57**, 5455 (2011).
- [84] 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 (2013).
- [85] K. Li, “Second-order asymptotics for quantum hypothesis testing,” *Ann. Stat.* **42**, 171 (2014).
- [86] N. Datta and F. Leditzky, “Second-Order Asymptotics for Source Coding, Dense Coding, and Pure-State Entanglement Conversions,” *IEEE Trans. Inf. Theory* **61**, 582 (2015).
- [87] P. Faist, T. Sagawa, K. Kato, H. Nagaoka, and F. G. S. L. Brandão, “Macroscopic Thermodynamic Reversibility in Quantum Many-Body Systems,” *Phys. Rev. Lett.* **123**, 250601 (2019).
- [88] T. Sagawa, P. Faist, K. Kato, K. Matsumoto, H. Nagaoka, and F. G. S. L. Brandao, “Asymptotic Reversibility of Thermal Operations for Interacting Quantum Spin Systems via Generalized Quantum Stein’s Lemma,” (2019), arXiv:1907.05650 [quant-ph] .
- [89] R. van der Meer, N. H. Y. Ng, and S. Wehner, “Smoothed generalized free energies for thermodynamics,” *Phys. Rev. A* **96**, 062135 (2017).
- [90] W. van Dam and P. Hayden, “Universal entanglement transformations without communication,” *Phys. Rev. A* **67**, 060302 (2003).

- [91] A. Anshu, V. K. Devabathini, and R. Jain, “Quantum communication using coherent rejection sampling,” *Phys. Rev. Lett.* **119**, 120506 (2017).
- [92] A. Anshu, M.-H. Hsieh, and R. Jain, “Quantifying Resources in General Resource Theory with Catalysts,” *Phys. Rev. Lett.* **121**, 190504 (2018).
- [93] P. Ćwikliński, M. Studziński, M. Horodecki, and J. Oppenheim, “Limitations on the evolution of quantum coherences: Towards fully quantum second laws of thermodynamics,” *Phys. Rev. Lett.* **115**, 210403 (2015).
- [94] M. Lostaglio, D. Jennings, and T. Rudolph, “Description of quantum coherence in thermodynamic processes requires constraints beyond free energy,” *Nat. Commun.* **6**, 6383 (2015).
- [95] M. Lostaglio, K. Korzekwa, D. Jennings, and T. Rudolph, “Quantum Coherence, Time-Translation Symmetry, and Thermodynamics,” *Phys. Rev. X* **5**, 021001 (2015).
- [96] K. Korzekwa, M. Lostaglio, J. Oppenheim, and D. Jennings, “The extraction of work from quantum coherence,” *New J. Phys.* **18**, 023045 (2016).
- [97] M. Perarnau-Llobet, E. Bäumer, K. V. Hovhannisyanyan, M. Huber, and A. Acin, “No-Go Theorem for the Characterization of Work Fluctuations in Coherent Quantum Systems,” *Phys. Rev. Lett.* **118**, 070601 (2017).
- [98] E. Bäumer, M. Lostaglio, M. Perarnau-Llobet, and R. Sampaio, “Fluctuating Work in Coherent Quantum Systems: Proposals and Limitations,” in *Thermodynamics in the Quantum Regime*, Fundamental Theories of Physics, Vol. 195, edited by B. F., C. L., G. C., A. J., and A. G. (Springer, Cham, 2018) pp. 275–300.
- [99] G. Gour, D. Jennings, F. Buscemi, R. Duan, and I. Marvian, “Quantum majorization and a complete set of entropic conditions for quantum thermodynamics,” *Nat. Commun.* **9**, 5352 (2018).
- [100] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, 2013).
- [101] F. Buscemi, “Degradable channels, less noisy channels, and quantum statistical morphisms: An equivalence relation,” *Problems of Information Transmission* **52**, 201 (2016).
- [102] F. Buscemi, D. Sutter, and M. Tomamichel, “An information-theoretic treatment of quantum dichotomies,” *Quantum* **3**, 209 (2019).
- [103] P. Shor, “Structure of Unital Maps and the Asymptotic Quantum Birkhoff Conjecture,” *Steklov Mathematical Institute Seminar* (2010).
- [104] U. Haagerup and M. Musat, “Factorization and Dilation Problems for Completely Positive Maps on von Neumann Algebras,” *Commun. Math. Phys.* **303**, 555 (2011).
- [105] U. Haagerup and M. Musat, “An Asymptotic Property of Factorizable Completely Positive Maps and the Connes Embedding Problem,” *Commun. Math. Phys.* **338**, 721 (2015).
- [106] C. B. Mendl and M. M. Wolf, “Unital quantum channels – convex structure and revivals of birkhoff’s theorem,” *Communications in Mathematical Physics* **289**, 1057 (2009).
- [107] A. Müller-Hermes and C. Perry, “All unital qubit channels are 4-noisy operations,” *Lett. Math. Phys.* **109**, 1 (2019).
- [108] P. Faist, J. Oppenheim, and R. Renner, “Gibbs-preserving maps outperform thermal operations in the quantum regime,” *New J. Phys.* **17**, 043003 (2015).



- [109] J. M. Renes, “Work cost of thermal operations in quantum thermodynamics,” *The European Physical Journal Plus* **129**, 153 (2014).
- [110] J. Åberg, “Truly work-like work extraction via a single-shot analysis,” *Nat. Commun.* **4**, 1925 (2013).
- [111] N. Brunner, N. Linden, S. Popescu, and P. Skrzypczyk, “Virtual qubits, virtual temperatures, and the foundations of thermodynamics,” *Phys. Rev. E* , 51117 (2011).
- [112] J. A. Vaccaro and S. M. Barnett, “Information erasure without an energy cost,” *Proc. R. Soc. A* **467**, 1770 (2011).
- [113] J. Åberg, “Catalytic coherence,” *Phys. Rev. Lett.* **113**, 150402 (2014).
- [114] M. P. Müller, “Correlating Thermal Machines and the Second Law at the Nanoscale,” *Phys. Rev. X* **8**, 041051 (2018).
- [115] M. Lostaglio, M. P. Müller, and M. Pastena, “Stochastic Independence as a Resource in Small-Scale Thermodynamics,” *Phys. Rev. Lett.* **115**, 150402 (2015).
- [116] M. P. Müller and M. Pastena, “A Generalization of Majorization that Characterizes Shannon Entropy,” *IEEE Trans. Inf. Theory* **62**, 1711 (2016).
- [117] W. H. Zurek, “Decoherence, einselection, and the quantum origins of the classical,” *Rev. Mod. Phys.* **75**, 715 (2003).
- [118] A. Ambainis, A. Ambainis, M. Mosca, A. Tapp, and R. De Wolf, “Private Quantum Channels,” *Proc. 41ST Annu. Symp. Found. Comput. Sci.* , 547 (2000).
- [119] J. Preskill, “Quantum Information, Chapter 3,” *Lecture Notes for Ph219/CS219* (Access:2020).
- [120] K. M. R. Audenaert and S. Scheel, “On random unitary channels,” *New J. Phys.* **10**, 023011 (2008).
- [121] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Yunger Halpern, “The resource theory of informational nonequilibrium in thermodynamics,” *Phys. Rep.* **583**, 1 (2015).
- [122] M. A. Nielsen and G. Vidal, “Majorization and the interconversion of bipartite states,” *Quantum Info. Comput.* **1**, 76 (2001).
- [123] M. A. Nielsen, “Conditions for a Class of Entanglement Transformations,” *Phys. Rev. Lett.* **83**, 436 (1999).
- [124] R. Orús, “A practical introduction to tensor networks: Matrix product states and projected entangled pair states,” *Annals of Physics* **349**, 117 (2014).
- [125] M. Klimesh, “Inequalities that collectively completely characterize the catalytic majorization relation,” (2007), arXiv:0709.3680 [quant-ph] .
- [126] S. Turgut, “Catalytic transformations for bipartite pure states,” *Journal of Physics A: Mathematical and Theoretical* **40**, 12185 (2007).
- [127] J. G. Richens, Á. M. Alhambra, and L. Masanes, “Finite-bath corrections to the second law of thermodynamics,” *Phys. Rev. E* **97**, 062132 (2018).

- [128] D. Reeb and M. M. Wolf, “An improved Landauer principle with finite-size corrections,” *New J. Phys.* **16**, 103011 (2014).
- [129] H. Wilming, R. Gallego, and J. Eisert, “Second law of thermodynamics under control restrictions,” *Phys. Rev. E* **93**, 042126 (2016).
- [130] C. Perry, P. Ćwikliński, J. Anders, M. Horodecki, and J. Oppenheim, “A sufficient set of experimentally implementable thermal operations for small systems,” *Phys. Rev. X* **8**, 041049 (2018).
- [131] M. Lostaglio, Á. M. Alhambra, and C. Perry, “Elementary Thermal Operations,” *Quantum* **2**, 52 (2016).
- [132] J. Eisert, M. Friesdorf, and C. Gogolin, “Quantum many-body systems out of equilibrium,” *Nat. Phys.* **11**, 124 (2015).
- [133] S. Daftuar and P. Hayden, “Quantum state transformations and the Schubert calculus,” *Ann. Phys. (N. Y.)* **315**, 80 (2005).
- [134] B. Groisman, S. Popescu, and A. Winter, “Quantum, classical, and total amount of correlations in a quantum state,” *Phys. Rev. A* **72**, 032317 (2005).
- [135] C. Majenz, M. Berta, F. Dupuis, R. Renner, and M. Christandl, “Catalytic Decoupling of Quantum Information,” *Phys. Rev. Lett.* **118**, 080503 (2017).
- [136] P. Hayden, D. Leung, P. W. Shor, and A. Winter, “Randomizing Quantum States: Constructions and Applications,” *Commun. Math. Phys.* **250**, 371 (2004).
- [137] C. Dankert, R. Cleve, J. Emerson, and E. Livine, “Exact and approximate unitary 2-designs and their application to fidelity estimation,” *Phys. Rev. A* **80**, 012304 (2009).
- [138] D. Gross, K. Audenaert, and J. Eisert, “Evenly distributed unitaries: On the structure of unitary designs,” *J. Math. Phys.* **48**, 052104 (2007).
- [139] H. Dale, D. Jennings, and T. Rudolph, “Provable quantum advantage in randomness processing,” *Nat. Commun.* **6**, 8203 (2015).
- [140] M. B. Hastings, “Random unitaries give quantum expanders,” *Phys. Rev. A* **76**, 032315 (2007).
- [141] A. W. Harrow, “Quantum expanders from any classical cayley graph expander,” *Quantum Info. Comput.* **8**, 715 (2008).
- [142] M. Lostaglio and M. P. Müller, “Coherence and Asymmetry Cannot be Broadcast,” *Phys. Rev. Lett.* **123**, 020403 (2019).
- [143] H. Wilming, M. P. Mueller, and P. Boes, “Catalytic Entropy Conjecture - Problem 45,” (2019).
- [144] I. Sergeev, *Equivalence of the catalytic entropy conjecture and the transitivity conjecture is plausible*, Master’s thesis, ETH Zurich (2019).
- [145] S. Rethinasamy and M. M. Wilde, “Relative entropy and catalytic relative majorization,” (2019), arXiv:1912.04254 [quant-ph] .
- [146] G. Crooks, “Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences,” *Phys. Rev. E* **60**, 2721 (1999).

- [147] M. Esposito, “Nonequilibrium fluctuations, fluctuation theorems, and counting statistics in quantum systems,” *Rev. Mod. Phys.* **81**, 1665 (2009).
- [148] M. Campisi, P. Hänggi, and P. Talkner, “Colloquium: Quantum fluctuation relations: Foundations and applications,” *Rev. Mod. Phys.* **83**, 771 (2011).
- [149] Á. M. Alhambra, L. Masanes, J. Oppenheim, and C. Perry, “Fluctuating Work: From Quantum Thermodynamical Identities to a Second Law Equality,” *Phys. Rev. X* **6**, 041017 (2016).
- [150] J. Åberg, “Fully Quantum Fluctuation Theorems,” *Phys. Rev. X* **8**, 011019 (2018).
- [151] T. Sagawa and M. Ueda, “Nonequilibrium thermodynamics of feedback control,” *Phys. Rev. E* **85**, 21104 (2012).
- [152] T. Sagawa and M. Ueda, “Minimal Energy Cost for Thermodynamic Information Processing: Measurement and Information Erasure,” *Phys. Rev. Lett.* **102**, 250602 (2009).
- [153] T. Sagawa and M. Ueda, “Second Law of Thermodynamics with Discrete Quantum Feedback Control,” *Phys. Rev. Lett.* **100**, 80403 (2008).
- [154] T. Sagawa and M. Ueda, “Fluctuation Theorem with Information Exchange: Role of Correlations in Stochastic Thermodynamics,” *Phys. Rev. Lett.* **109**, 180602 (2012).
- [155] M. Esposito, K. Lindenberg, and C. Van den Broeck, “Entropy production as correlation between system and reservoir,” *New J. Phys.* **12**, 13013 (2010).
- [156] F. Sapienza, F. Cerisola, and A. J. Roncaglia, “Correlations as a resource in quantum thermodynamics,” *Nat. Commun.* **10**, 2492 (2019).
- [157] F. Bakhshinezhad, F. Clivaz, G. Vitagliano, P. Erker, A. Rezakhani, M. Huber, and N. Friis, “Thermodynamically optimal creation of correlations,” *J. Phys. A Math. Theor.* **52**, 465303 (2019).
- [158] E. Jaynes, “Information theory and statistical mechanics,” *Phys. Rev.* **106**, 620 (1957).
- [159] E. Jaynes, “Information theory and statistical mechanics. II,” *Phys. Rev.* **108**, 171 (1957).
- [160] G. Vidal, “Entanglement Renormalization,” *Phys. Rev. Lett.* **99**, 220405 (2007).
- [161] G. Vidal, “Class of Quantum Many-Body States That Can Be Efficiently Simulated,” *Phys. Rev. Lett.* **101**, 110501 (2008).



## BACK MATTER

---

### 9.1 ZUSAMMENFASSUNG

Quantenthermodynamik ist ein relativ junges und wachsendes Forschungsfeld, in dem Methoden und Konzepte aus der Quanteninformationstheorie verwendet werden um die Gesetzmäßigkeiten der Thermodynamik für kleine Quanten-Systeme zu studieren, auf die der thermodynamische Grenzfall nicht notwendigerweise Anwendung findet. Sie trägt hiermit bei sowohl zu den Grundlagen der Thermodynamik, indem sie die Gesetze der phänomenologischen Thermodynamik konstruktiv und “bottom-up” auf der Basis der Postulate der Quantenmechanik herleitet, als auch zur Praxis, indem sie die theoretischen Grundlagen für eine immer größere Menge an Experimenten liefert, in denen thermodynamische Maschinen gebaut werden, für deren adequate Beschreibung Quantenmechanik vonnöten ist.

Ein zentrales mathematisches Werkzeug der Quantenthermodynamik sind sogenannte Ressourcentheorien, insbesondere die Ressourcentheorie der *thermal operations*, in welcher die thermodynamische Wechselwirkung eines Systems mit einem Wärmebad und weiteren Systemen wie Batterien, Uhren und Katalysatoren modelliert wird. In dieser kumulativen Dissertation werden verschiedene Erweiterungen und Modifikationen dieses Modells eingeführt, auf deren Basis neue Ergebnisse bezüglich der möglichen thermodynamischen Evolution von Quantensystemen hergeleitet werden. Konkret werden, nach einer systematischen Einführung in die Ressourcentheorie der *thermal operations*, Antworten auf die folgenden Fragen entwickelt: i) Wie groß muss ein Wärmebad mindestens sein, damit ein System einen gegebenen stochastischen oder thermodynamischen Prozess durchlaufen kann? ii) Hängt die Antwort auf Frage i) davon ab, ob die Interaktion zwischen Bad und System quantisch oder klassisch ist? iii) Welche thermodynamischen Zustandsänderungen eines System sind möglich in der Wechselwirkung mit katalytischen Hilfs-Systemen? iv) Welche Zustandsübergänge können operational erwirkt werden, wenn die zugrundeliegenden Mikrozustände von Bad und System nur teilweise bekannt sind? v) Wie können wir den empirischen Erfolg des kanonischen Ensembles als Beschreibung von Systemen im thermodynamischen Gleichgewicht verstehen?

Diese Fragen mögen sehr unterschiedlich erscheinen, aber es wird klar werden, dass sie alle mithilfe der gleichen mathematischen Werkzeuge behandelt werden können. Somit sind die Ergebnisse, die hier präsentiert werden, nicht nur interessant für sich genommen – es wird etwa gezeigt, dass quantische Zufallsprozesse strikt mächtiger sein können als ihre klassischen Gegenstücke existieren, oder dass mithilfe von Katalysatoren extensive Mengen an Arbeit von makroskopischen Systemen im thermischen Gleichgewicht mit nicht-verschwindender Wahrscheinlichkeit extrahiert werden können –, sondern sie illustrieren auch die Breite der Anwendbarkeit der Methoden der Quantenthermodynamik als Bindeglied zwischen Quantenmechanik und Thermodynamik.

## 9.2 ANTEILE DES AUTORS AN ZUGRUNDELIEGENDEN ARBEITEN

Im Folgenden wird für jede der Publikationen, die im Verlauf dieser Dissertation entstanden sind, der Anteil des Verfassers bei Konzeption, Durchführung und Verfassung der Publikation aufgeführt.

- [1] Der Verfasser war federführend in diesem Projekt. Er hat wesentliche Teile zur Konzeption, Herleitung der Ergebnisse sowie Verfassung der Publikation beigetragen.
- [2] Der Verfasser ist hauptverantwortlich für die Konzeption des Projektes und hat wichtige Beiträge zur Herleitung der Ergebnisse geleistet. Er hat wesentliche Teile zur Konzeption, Herleitung der Ergebnisse sowie Verfassung der Publikation beigetragen.
- [3] Der Verfasser war maßgeblich an Konzeption des Projekts, der Herleitung der Hauptergebnisse, sowie der Formulierung der Publikation beteiligt.
- [4] Der Verfasser war federführend in diesem Projekt. Er hat wesentliche Teile zur Herleitung der Ergebnisse sowie zur Formulierung der Publikation beigetragen.
- [5] Der Verfasser hat substanzielle Beiträge zur Herleitung der Ergebnisse sowie zur Formulierung der Publikation beigetragen.
- [6] Der Verfasser war federführend in diesem Projekt. Er hat wesentliche Teile zur Herleitung der Ergebnisse sowie Formulierung der Publikation beigetragen.
- [7] Der Verfasser war maßgeblich an Konzeption des Projekts, der Herleitung der Hauptergebnisse, sowie der Formulierung der Publikation beteiligt.
- [8] Der Verfasser war federführend in diesem Projekt. Er hat wesentliche Teile zur Konzeption, Herleitung der Ergebnisse sowie zur Formulierung der Publikation beigetragen.
- [9] Der Verfasser war maßgeblich an Konzeption des Projekts, der Herleitung der Hauptergebnisse, sowie der Formulierung der Publikation beteiligt.

## 9.3 SELBSTSTÄNDIGKEITSERKLÄRUNG

Ich versichere hiermit, dass diese Arbeit von niemand anderem als meiner Person verfasst worden ist. Alle verwendeten Hilfsmittel wie Berichte, Bücher, Internetseiten oder ähnliches sind im Literaturverzeichnis angegeben, Zitate aus fremden Arbeiten sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungskommission vorgelegt und auch nicht veröffentlicht.

Paul Boës