

The resource theory of informational nonequilibrium in thermodynamics

Gilad Gour,^{1,2,*} Markus P. Müller,^{3,4,†} Varun Narasimhachar,^{1,5,‡} Robert W. Spekkens,^{3,§} and Nicole Yunger Halpern^{3,6,¶}

¹*Institute for Quantum Science and Technology, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4*

²*Department of Mathematics and Statistics, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4*

³*Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario Canada N2L 2Y5*

⁴*Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 19, D-69120 Heidelberg, Germany*

⁵*Department of Physics and Astronomy, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4*

⁶*Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, CA 91125, USA.*

(Dated: Sept. 24, 2013; revised April 13, 2015)

We review recent work on the foundations of thermodynamics in the light of quantum information theory. We adopt a resource-theoretic perspective, wherein thermodynamics is formulated as a theory of what agents can achieve under a particular restriction, namely, that the only state preparations and transformations that they can implement for free are those that are thermal at some fixed temperature. States that are out of thermal equilibrium are the resources. We consider the special case of this theory wherein all systems have trivial Hamiltonians (that is, all of their energy levels are degenerate). In this case, the only free operations are those that add noise to the system (or implement a reversible evolution) and the only nonequilibrium states are states of *informational nonequilibrium*, that is, states that deviate from the maximally mixed state. The degree of this deviation we call the state's *nonuniformity*; it is the resource of interest here, the fuel that is consumed, for instance, in an erasure operation. We consider the different types of state conversion: exact and approximate, single-shot and asymptotic, catalytic and noncatalytic. In each case, we present the necessary and sufficient conditions for the conversion to be possible for any pair of states, emphasizing a geometrical representation of the conditions in terms of Lorenz curves. We also review the problem of quantifying the nonuniformity of a state, in particular through the use of generalized entropies, and that of quantifying the gap between the nonuniformity one must expend to achieve a single-shot state preparation or state conversion and the nonuniformity one can extract in the reverse operation. Quantum state conversion problems in this resource theory can be shown to be always reducible to their classical counterparts, so that there are no inherently quantum-mechanical features arising in such problems. This body of work also demonstrates that the standard formulation of the second law of thermodynamics is inadequate as a criterion for deciding whether or not a given state transition is possible.

Contents

I. Introduction	2	A. Nonuniformity monotones built from convex functions on the reals	14
II. Preliminaries	5	1. Schur-convexity	14
A. Free operations, monotones, and witnesses	5	2. Schur-convexity relative to a distribution q	15
B. Reducing quantum state conversion problems to their classical counterparts	6	3. Schur-convexity relative to the uniform state	15
III. Quasi-order of states under noisy operations	7	4. Schur-convexity-preserving functions	16
A. Equivalence classes under noisy operations	7	5. Examples, including order- p Rényi nonuniformities	16
B. Deterministic interconversion of nonuniform states	10	B. Nonuniformity monotones arising from the geometry of the Lorenz curve	20
1. The result	10	C. Nonuniformity monotones from other Schur-convex functions	21
2. Some consequences	11	V. Exact state conversion	22
3. Implementation	12	A. A standard form of nonuniformity: sharp states	22
IV. Nonuniformity monotones	13	B. Nonuniformity of formation and distillable nonuniformity	23
		C. Nonuniformity cost and yield of state conversion	25
		D. Catalysis	28
		1. Quasi-order of states under noisy operations assisted by a catalyst	29
		2. Nonuniformity of formation and distillable nonuniformity in the presence of a catalyst	31
		3. Nonuniformity cost and yield of catalytic state conversion	32
		E. Inadequacy of the second law as a criterion for state conversion	32

*Electronic address: gour@ucalgary.ca

†Electronic address: markus@mpmueller.net

‡Electronic address: vnasim@ucalgary.ca

§Electronic address: rspekkens@perimeterinstitute.ca

¶Electronic address: NicoleYH@caltech.edu

VI. Approximate state conversion	33
A. Metrics over state space	33
B. Smoothed entropies	35
C. Nonuniformity of formation and distillable nonuniformity for approximate state preparation	36
D. Asymptotic state conversion	37
E. Nonuniformity cost and yield of asymptotic state conversion	39
F. Approximate catalysis	40
VII. Conclusions	41
VIII. Acknowledgements	43
A. Proofs	43
B. Strong noisy-trumping and $p < 0$ Rényi nonuniformities	47
C. Approximate distillation	49
References	50

I. INTRODUCTION

A resource theory is defined by identifying a set of experimental operations that are considered free, that is, a set which can be used without limit, while the rest are considered expensive and are thus treated as resources (Coecke *et al.*, 2014). A *quantum* resource theory is defined by identifying a set of free operations within the set of all quantum operations. For instance, the restriction to local quantum operations and classical communication defines the resource theory of entanglement (Horodecki *et al.*, 2009), the restriction to symmetric quantum operations defines the resource theory of asymmetry (Bartlett *et al.*, 2007; Gour and Spekkens, 2008; Marvian and Spekkens, 2013), and the restriction to stabilizer operations defines a theory of quantum computational resources (Veitch *et al.*, 2014).

The resource theory that is the subject of this article is defined by the following set of free quantum operations:

- one can implement any unitary on a system,
- one can prepare any system in the uniform quantum state (i.e., the completely mixed state),
- one can take a partial trace over any subsystem of a system.

In the most general such operation, an ancilla prepared in the uniform state is adjoined to the system of interest, they are coupled unitarily, and finally one takes a partial trace over some subsystem. These have been called *noisy operations* in (Horodecki *et al.*, 2003a,b), where this resource theory was first studied¹. If one comes to possess a system that is *not* in the uniform state, then because

¹ They have also been called *exactly factorizable maps* by Shor (Shor, 2010); see also (Haagerup and Musat, 2011)

such a state cannot be prepared for free, it is a resource. We call such states *nonuniform*, and we call the aspects of a state that are relevant to the resource theory the state’s *nonuniformity properties*.

Nonuniform states are also resources insofar as they can be useful for simulating operations that are outside the noisy class: if a certain process taking a given initial state to a desired final state is not achievable via noisy operations, the process can always be rendered possible by adding an appropriate nonuniform state to the given initial state and suffering the consumption (partial or complete) of this added resource. An example of a process that would normally be disallowed under noisy operations, but becomes possible by supplying (and using up) an added nonuniformity resource, is an erasure operation. Erasure is “informational work”, and therefore one can understand the nonuniformity of a state as the potential for doing informational work.

The resource theory of nonuniformity is interesting primarily because of the light that it sheds on another, more intricate and more relevant, resource theory: the one where the free operations are those that are thermal relative to some temperature T , and the resources are states that are not thermal at this temperature, or *athermal states* (Brandão *et al.*, 2015, 2013; Horodecki and Oppenheim, 2013; Janzing *et al.*, 2000; Skrzypczyk *et al.*, 2013, 2014).²

The thermal operations relative to a background temperature T are defined by the following capabilities: One can prepare any subsystem in its thermal state at temperature T , one can implement any unitary that commutes with the total Hamiltonian³, and one can trace over the subsystems. It follows that the resource theory of nonuniformity is simply a special case of the resource theory of athermality (for any temperature) where all systems of interest have trivial Hamiltonians (i.e., all their energy levels are degenerate).

To achieve a proper conceptual understanding of thermodynamics, it is critical to disentangle those aspects of the theory that are due to considerations of energetics and those that are due to considerations of information theory. Indeed, there is now a large literature on the role of information theory in thermodynamics, centered around such topics as Maxwell’s demon (Maruyama *et al.*, 2009), Szilard’s engine (Szilard, 1929), the thermodynamic reversibility of computation and Landauer’s principle (Bennett, 1973, 1982; Landauer, 1961), and the use of maximum entropy principles in statistical mechan-

² The resource-theoretic approach to thermodynamics is cognate with the earlier work of Lieb and Yngvason (Lieb and Yngvason, 1999), where classical thermodynamics is formulated in terms of interconvertibility relations between equilibrium states.

³ It is presumed that the couplings between subsystems can be entirely controlled by the experimenter, so that in the absence of any such intervention, the overall Hamiltonian is just a sum of the free Hamiltonians for each subsystem.

ics (Jaynes, 1957a,b). (See also (Leff and Rex, 2003).) By focussing on the resource theory of athermality for the special case of systems that are energy-degenerate, we can begin to understand what aspects of thermodynamics are purely informational. In essence, we are here studying the particular type of thermal nonequilibrium corresponding to purely *informational nonequilibrium*. The term ‘nonuniformity’ can be understood as a shorthand for ‘informational nonequilibrium’.

In addition, many results in athermality theory can be inferred from results in nonuniformity theory. As such, developing the latter can help to answer questions in the former. Another motivation for focussing on the case of trivial Hamiltonians is that in building up one’s understanding of a field, it is always useful to start with the simplest case.

In this article, we focus on determining necessary and sufficient conditions for it to be possible to convert one state to another under noisy operations—exactly and approximately, single-shot and asymptotically, with and without a catalyst—and on finding measures of nonuniformity.

We demonstrate that the answers to these sorts of questions in the quantum case can be inferred from their answers in the classical resource theory of nonuniformity, which is defined in precise analogy with the quantum theory, as follows. A classical system is represented by a physical state space, or equivalently by a random variable, the valuations of which correspond to the physical states (i.e. the points in that space). The states of such a system that are of interest in the resource theory (and which are the analogues of the quantum states) are *statistical states*, that is, probability distributions over the physical state space. The free operations are:

- one can implement any permutation on a system’s physical state space,
- one can prepare any system in the uniform statistical state (i.e. the uniform probability distribution),
- one can marginalize over any subsystem of a system.

The fact that state conversion problems in the quantum resource theory of nonuniformity reduce to their classical counterparts shows that there is nothing inherently quantum in any problem of state conversion in this theory.

This article seeks to present what is currently known about the resource theory of nonuniformity in a systematic and pedagogical fashion. In particular, we characterize the equivalence classes of states under noisy operations by *Lorenz curves*, thereby providing a geometric perspective on the problem of defining measures of nonuniformity and state conversion problems, a perspective that makes the proofs of certain results more straightforward and intuitive.

We draw on various sources.

The resource-theoretic approach to thermodynamics has seen a great deal of activity in the past few years, primarily by researchers in the field of quantum information (Åberg, 2013; Allahverdyan *et al.*, 2004; Brandão *et al.*, 2015, 2013; Dahlsten *et al.*, 2009; Del Rio *et al.*, 2011; Egloff *et al.*, 2012; Faist *et al.*, 2012; Horodecki and Oppenheim, 2013; Janzing *et al.*, 2000; Linden *et al.*, 2010; Skrzypczyk *et al.*, 2013, 2014). Many of the results we present are rederivations of results from these works, or special cases thereof. In this sense, our article provides a review of some of this literature.

Insofar as such work (and this article) does not confine its attention to macroscopic thermodynamic variables, but rather focusses on the quantum state of the microscopic degrees of freedom, it is perhaps best described as a resource-theoretic approach to *statistical mechanics*. Because it also does not confine its attention to the family of states that are thermal for some temperature, it concerns *nonequilibrium* statistical mechanics.⁴ Finally, because it seeks to understand not just what occurs *on average* but what happens in a single shot, in particular, in the worst case, it is an application to statistical mechanics of the ideas of single-shot information theory.⁵

Because of the reduction of quantum questions about state conversion to their classical counterparts, there are a number of works on *classical* statistical mechanics that can be brought to bear on our problem. In particular, our analysis is informed by the work of Ernst Ruch and his collaborators, who adopted a kind of resource-theoretic perspective on state conversion problems already in the 1970s (Ruch, 1975a,b; Ruch and Lesche, 1978; Ruch and Mead, 1976; Ruch *et al.*, 1978).

We also make heavy use of the mathematical literature on majorization theory (Bhatia, 1997; Hardy *et al.*, 1952; Horn and Johnson, 2012; Joe, 1990), in particular, the canonical text on the subject by Marshall, Olkin and Arnold (Marshall *et al.*, 2010), which we shall refer to throughout the article as simply MOA.

Finally, in various places (in particular in our discussion of measures of nonuniformity) we make use of work concerning the zoo of generalized entropies (Datta, 2009; van Erven and Harremoës, 2010; Gorbun *et al.*, 2010; Renner, 2008; Renner and Wolf, 2004).

⁴ It is important to distinguish between two notions of nonequilibrium here. In the resource theory of athermality, there is always a background temperature T defining which states are free, and any state that is not *thermal at temperature T* is considered a resource. In this context, if a state is thermal for a temperature T' differing from T , it is natural to describe it as being *out of equilibrium*. Conventionally, however, we use the term *equilibrium statistical mechanics* to refer to situations wherein every system is in a thermal state, but where different systems may be at different temperatures. In this context, *nonequilibrium* means not thermal relative to *any* temperature.

⁵ And insofar as the *thermodynamic limit* typically involves such averaging procedures, it is concerned with thermodynamic questions outside of this limit.

We now review the structure of the article.

Section II covers preliminary material. In particular, we clarify the definition of noisy quantum operations, nonuniformity monotones and conversion witnesses, and we explain why state conversion problems in the quantum resource theory of nonuniformity can be reduced to their classical counterparts.

In Section III, we characterize the quasi-order of states under noisy operations. We begin by determining the equivalence classes of states under noisy operations. We show that these can be associated with a mathematical object known as the Lorenz curve of the state. We then present the necessary and sufficient conditions for one state to be deterministically converted to another by noisy operations in terms of the Lorenz curves of the states.

In Section IV, we discuss various techniques for constructing nonuniformity monotones. We show that various generalized entropies and relative entropies lead to useful monotones, some of which have operational interpretations. We also show how certain monotones can be inferred from the geometry of the Lorenz curve.

We consider various kinds of exact state conversion in Section V. We begin by defining a standard form of the resource of nonuniformity in terms of a one-parameter family of states which we term the *sharp states*. In terms of these sharp states, we can determine the amount of nonuniformity one requires to form a given state, as well as the amount of nonuniformity that one can distill from that state. These are shown to be related to nonuniformity monotones based on the Rényi entropies of order 0 and order ∞ . We also consider the amount of nonuniformity that must be paid to achieve a given state conversion or which can be recovered in addition to achieving the state conversion. We apply these results to determine some simple sufficient conditions for state conversion. Next, we consider exact state conversion in the presence of a catalyst, that is, a nonuniform state which can be used in the conversion process but which must be returned intact at the end of the protocol. In particular, we note the presence of nontrivial catalysis in our resource theory, but also the uselessness of sharp states as catalysts. We review the necessary and sufficient conditions for a state conversion to be possible by catalysis. Finally, we discuss the inadequacy of the standard formulation of the second law of thermodynamics (nondecrease of entropy) as a criterion for deciding whether a state conversion is possible, with or without a catalyst, and we discuss the criteria that actually do the job.

In Section VI, we shift our attention to approximate state conversion. We begin by formalizing the notion of approximate state conversion in terms of a state being mapped to one that is ε -close to another relative to some contractive metric on the state space, and discuss subtleties of the quantum to classical reduction. We then review the notion of smoothed entropies. We reconsider the problem of determining the nonuniformity of formation and the distillable nonuniformity when the state

conversion is allowed to be approximate, and we show that the answers are provided by the smoothed versions of the entropies that characterize exact conversion. We then use these results to provide a simple proof of the rate of asymptotic conversion between states, that is, the rate at which one state can be converted approximately to another in the limit of arbitrarily many copies. The asymptotic results are in turn applied to determine the nonuniformity cost and yield of approximate state conversion when one has arbitrarily many copies to convert. Finally, we consider various notions of approximate state conversion in the presence of a catalyst.

In Section VII, we discuss a few open problems and highlight some of the overarching conclusions of the article.

This article provides a review and synthesis of many known results, but along the way we also present a number of novel results, which we now summarize.

We demonstrate that all exact state conversion problems in the quantum resource theory of nonuniformity reduce to the corresponding problems in the classical resource theory, and that all *approximate* state conversion problems can also be reduced to their classical counterparts as long as one makes a judicious choice of the metric over states by which one judges the degree of approximation. We generalize many known results on state conversion and on majorization to the case where the input system and the output system are of different dimensions. In particular, we describe a general scheme for building nonuniformity monotones from Schur-convex functions in this case. We introduce the notion of a *state conversion witness*: a function of a pair of states in terms of which one can specify a necessary condition for the possibility of conversion from one state to the other, or a necessary condition for the impossibility of such a conversion, or both. We demonstrate how such witnesses are the appropriate tool for summarizing some known results, and are as such more versatile than monotones. We introduce a class of states, which we call the *sharp states*, that serve as a gold standard form of nonuniformity and that naturally generalize the notion of ‘ n pure bits’ to the case that n is not an integer. We use sharp states to quantify the nonuniformity of formation (or of distillation) of an arbitrary state and the nonuniformity cost (or yield) of an arbitrary state conversion. In the case of approximate conversion, we prove exact expressions for the nonuniformity of formation and of distillation for a class of judiciously-chosen metrics over the states. For exact conversion using a catalyst, we note the uselessness of sharp states as catalysts and its implication that having access to an ideal measurement cannot catalyze any state conversion. Finally, we present a new and simplified proof of the rate of asymptotic conversion between nonuniform states.

II. PRELIMINARIES

A. Free operations, monotones, and witnesses

We begin by formalizing the definition of a noisy quantum operation. A quantum operation is a completely positive trace-preserving linear map $\mathcal{E} : \mathcal{L}(\mathcal{H}_{\text{in}}) \rightarrow \mathcal{L}(\mathcal{H}_{\text{out}})$, where $\mathcal{L}(\mathcal{H})$ is the set of linear bounded operators acting on \mathcal{H} . In this paper, all Hilbert spaces are assumed to be finite-dimensional.

Definition 1 (Noisy quantum operations) *A noisy quantum operation \mathcal{E} is one that admits of the following decomposition.⁶ There must exist a finite-dimensional ancilla space \mathcal{H}_a and a unitary U on $\mathcal{H}_{\text{in}} \otimes \mathcal{H}_a$ such that, for all input states ρ_{in} ,*

$$\mathcal{E}(\rho_{\text{in}}) = \text{Tr}_{a'} \left(U \left(\rho_{\text{in}} \otimes \frac{1}{d_a} I_a \right) U^\dagger \right), \quad (2.1)$$

where $\mathcal{H}_{a'}$ is the space complementary to \mathcal{H}_{out} in the total Hilbert space, that is, $\mathcal{H}_{\text{out}} \otimes \mathcal{H}_{a'} = \mathcal{H}_{\text{in}} \otimes \mathcal{H}_a$, and where $d_a = \dim(\mathcal{H}_a)$.

Although at first glance, “degree of purity” might appear to be a good name for what we are calling “nonuniformity”, at second glance one recognizes that it is not because standard usage of the term “purity” takes a pure state of a 2-level system and a pure state of a 3-level to have *equal* degrees of purity, while the latter is a stronger resource than the former in the resource theory we are considering. The fact that these two pure states are not equivalent under noisy operations is counterintuitive for many quantum information theorists because it contrasts with the situation in entanglement theory, the resource theory with which they are most familiar. The difference arises because the free operations in entanglement theory (local operations and classical communication) allow one to prepare pure product states for free, thereby allowing a given state to be embedded into an arbitrarily large space, whereas noisy operations do not allow one to prepare pure states for free. More generally, two states with identical structure on their support have different amounts of resourcefulness depending on the dimension

of the space in which the state is embedded.⁷ The resourcefulness of a state under noisy operations is therefore not determined by its proximity to a pure state, but rather its distance from the uniform state, and therefore the term “nonuniformity” is more appropriate than the term “purity” as a description of the resource.

An important question in any resource theory is how to quantify the resource. A *resource monotone* is a function f from states to the real numbers such that if $\rho \mapsto \sigma$ by the free operations, then $f(\rho) \geq f(\sigma)$. For the set of operations given by noisy operations, we refer to the resource monotones as nonuniformity monotones:

Definition 2 (Quantum nonuniformity monotone) *A function M , mapping density operators to real numbers, is a nonuniformity monotone if for any two states ρ and σ (possibly of different dimensions), $\rho \mapsto \sigma$ by noisy operations implies that $M(\rho) \geq M(\sigma)$.*

Any measure of nonuniformity must at least be a nonuniformity monotone. When quantifying the relative resourcefulness of different states, there is a strong compulsion to think that the researcher’s task is to find the “one true measure to rule them all”. This tendency stems from an implicit assumption that any property worth quantifying must necessarily be totally ordered. In fact, it is more common to find that resource states only form a quasi-order, and no individual scalar measure can capture a quasi-order. So the search for the one true measure is in vain. The “total-order fallacy” must be resisted.

Nonetheless, resource monotones still have an important role to play. A *set* of such monotones *can* capture the quasi-order, in which case they are called a *complete set* of monotones. Also, if we define an operational task that requires the resource, it is then a well-defined question which resource monotone accurately quantifies the degree of success achievable in the task (according to some figure of merit) for a given state. The states become totally ordered relative to the task.

In this article, we describe some methods for generating a large number of nontrivial nonuniformity monotones, and we describe operational interpretations for some of these. We also identify some complete sets of nonuniformity monotones.

Besides quantifying nonuniformity, we will be primarily interested in problems of state conversion. A useful tool for characterizing such problems is the concept of a *state conversion witness*.

⁶ For a fixed pair of Hilbert spaces $\mathcal{H}_{\text{in}}, \mathcal{H}_{\text{out}}$, the set of maps from \mathcal{H}_{in} to \mathcal{H}_{out} of the form (2.1) is not in general topologically closed (Shor, 2010). Thus, strictly speaking, we define noisy quantum operations as those linear maps that can be *arbitrarily well approximated* in operator norm by maps of the form (2.1). This reflects the intuition that there is no physical difference between exact and arbitrarily accurate implementation of any map. Consequently, the set of noisy quantum operations becomes topologically closed. This allows us, in the following, to prove noisiness of quantum operations by showing how they can be approximated by maps of the form (2.1). Our noisy quantum operations are what (Shor, 2010) calls *strongly factorizable maps*.

⁷ This inequivalence can also be made slightly more intuitive by noting that it holds also in the resource theory of athermality and accounts for the fact that a Szilard engine at temperature T extracts different amounts of work from the two states: if it operates on a 2-level system that is prepared in a pure initial state, it can extract an expected work of $k_B T \log 2$, whereas if it operates on a 3-level system that is prepared in a pure initial state, it can extract $k_B T \log 3$.

Definition 3 (Quantum state conversion witness)

Let W be a real-valued function on pairs of quantum states, ρ and σ (possibly of different dimensions). W is said to be a go witness if $W(\rho, \sigma) \geq 0$ implies that $\rho \mapsto \sigma$ under the free quantum operations. W is said to be a no-go witness if $W(\rho, \sigma) < 0$ implies that $\rho \not\mapsto \sigma$ under the free quantum operations. Finally, W is said to be a complete witness if it is both a go witness and a no-go witness.

No-go witnesses were previously introduced under the name of *relative monotones* (Sanders and Gour, 2009).

Any resource monotone M defines a no-go witness by $W(\rho, \sigma) := M(\rho) - M(\sigma)$ because $W(\rho, \sigma) < 0$ implies $M(\sigma) > M(\rho)$, which implies that $\rho \not\mapsto \sigma$. It follows that M also defines a *resource witness* – that is, a function w such that $w(\sigma) < 0$ implies that σ is a resource – by $w(\sigma) := W(\rho_{\text{free}}, \sigma) = M(\rho_{\text{free}}) - M(\sigma)$, where ρ_{free} is any free state. A complete set $\{M_k\}$ of monotones defines a complete witness by $W(\rho, \sigma) := \min_k (M_k(\rho) - M_k(\sigma))$.

B. Reducing quantum state conversion problems to their classical counterparts

To define the *classical* resource theory of nonuniformity, we must characterize the set of noisy *classical* operations. Just as we have confined our attention to finite-dimensional Hilbert spaces in the quantum case, we confine our attention to discrete variables in the classical case (that is, finite information-carrying capacity in both cases). Let Ω_{in} be the discrete physical state space of the input system and Ω_{out} be the discrete physical state space of the output system. Let $S(\Omega)$ be the simplex of probability distributions over Ω , and let $V(\Omega)$ be the smallest linear vector space in which $S(\Omega)$ can be embedded. Clearly, every probability distribution on a sample space Ω can be represented by a vector in $V(\Omega)$. We call each such probability distribution a *state*. When we say that states are *normalized*, we refer to the fact that their components sum to total probability one.

The classical analogue of a quantum operation that is completely positive and trace-preserving is a classical operation which preserves positivity and normalization, hence taking probability distributions to probability distributions. For finite dimension, this is represented by a stochastic matrix, that is, a matrix D whose entries are real, are nonnegative, and satisfy $\sum_j D_{jk} = 1$.

We can now provide a definition of the free operations.

Definition 4 (Noisy classical operations) A *noisy classical operation* is a positivity-preserving and normalization-preserving map $D : V_{\text{in}} \rightarrow V_{\text{out}}$ that admits of the following decomposition.⁸ There must

exist an ancilla system with a discrete physical state space Ω_a and a permutation on $\Omega_{\text{in}} \times \Omega_a$ with an induced representation π on $V(\Omega_{\text{in}}) \otimes V(\Omega_a)$ such that, for all input states x_{in} ,

$$Dx_{\text{in}} = \sum_{\Omega_{a'}} \pi(x_{\text{in}} \otimes m_a), \quad (2.2)$$

where m_a is the normalized uniform distribution on Ω_a and $\Omega_{a'}$ is the physical state space complementary to Ω_{out} , that is, $\Omega_{\text{out}} \times \Omega_{a'} = \Omega_{\text{in}} \times \Omega_a$.

We note an important difference between the structure of the set of noisy quantum operations and the set of noisy classical operations.

Recall that a *unital* operation $\mathcal{E} : \mathcal{L}(\mathcal{H}_{\text{in}}) \rightarrow \mathcal{L}(\mathcal{H}_{\text{out}})$ is one for which $\mathcal{E}(\mathbb{1}_{\text{in}}/d_{\text{in}}) = \mathbb{1}_{\text{out}}/d_{\text{out}}$, where $\mathbb{1}_{\text{in}}$ and $\mathbb{1}_{\text{out}}$ are the identity operators on \mathcal{H}_{in} and \mathcal{H}_{out} , respectively (Mendl and Wolf, 2009).

Lemma 5 *Noisy quantum operations form a strict subset of the unital operations, and, in the case of equal dimension of input and output space, a strict superset of the mixtures of unitaries.*

The proof is described in Appendix A.

To compare with the classical case, we must specify the classical analogues of each of the classes of operations appearing in Lemma 5.

The classical analogue of a unital quantum operation is a stochastic matrix D that takes the uniform distribution on the input system to the uniform distribution on the output system, hence $\sum_k D_{jk} = d_{\text{in}}/d_{\text{out}}$ where d_{in} (d_{out}) is the dimension of the input (output) vector space. We will call such a stochastic matrix *uniform-preserving*. Note that, if the input and output spaces are of equal dimension, a uniform-preserving stochastic matrix D satisfies $\sum_j D_{jk} = \sum_k D_{jk} = 1$, in which case it is said to be doubly-stochastic. Finally, the classical analogue of a mixture of unitaries is a mixture of permutations of the physical state space.

The structure of the set of noisy classical operations is much more straightforward than the quantum one.

Lemma 6 *The set of noisy classical operations coincides with the set of uniform-preserving stochastic matrices and, in the case of equal dimension of input and output spaces (where the uniform-preserving stochastic matrices are the doubly-stochastic matrices), it coincides with the set of mixtures of permutations.*

set of classical operations of the form (2.2) between fixed vector spaces V_{in} and V_{out} is not topologically closed. This follows from the simple observation that input vectors x_{in} with all rational entries are mapped to output vectors Dx_{in} with all rational entries. Thus, as in the quantum case, we define noisy classical operations from V_{in} to V_{out} as those linear maps that can be *arbitrarily well approximated* in operator norm by maps of the form (2.2).

⁸ Analogously to the definition of noisy quantum operations, the

Here also, we relegate the proof to Appendix A. Note that the fact that every doubly-stochastic matrix can be written as a convex combination of permutations is known as Birkhoff's theorem (Birkhoff, 1946).

Because all unitaries are free operations in the quantum resource theory of nonuniformity, the only feature of a quantum state that is relevant for its nonuniformity properties is the vector of its eigenvalues (where an eigenvalue appears multiple times in the vector if it is degenerate and we explicitly include any zero eigenvalues). For any state ρ , we denote the vector of eigenvalues, listed in non-increasing order, by $\lambda(\rho)$. Therefore, the condition under which we can deterministically convert ρ to σ by noisy operations must be expressible in terms of $\lambda(\rho)$ and $\lambda(\sigma)$.

We can now state the reduction of the quantum state conversion problem to the corresponding classical problem, leaving the proof to Appendix A.

Lemma 7 *There exists a noisy quantum operation that achieves the quantum state conversion $\rho \mapsto \sigma$ if and only if there is a noisy classical operation that achieves the classical state conversion $\lambda(\rho) \mapsto \lambda(\sigma)$.*

Lemma 7 implies that all questions about exact state interconversion in the quantum resource theory of nonuniformity are answered by the classical theory, so studying the latter is sufficient.⁹ Later on (in Lemma 55), we will prove an analogous statement for approximate conversion.

In particular, Lemma 7 implies a reduction of quantum nonuniformity monotones and quantum state conversion witnesses to their classical counterparts, which are defined analogously to the quantum notions, Definitions 2 and 3, but with quantum states replaced by probability distributions, and noisy quantum operations replaced by noisy classical operations.

Corollary 8 *Consider M , a real-valued function over probability distributions, and M' , a real-valued function over quantum states, such that $M'(\rho) := M(\lambda(\rho))$, where $\lambda(\rho)$ is the vector of eigenvalues of ρ . Then M' is a quantum nonuniformity monotone if and only if M is a classical nonuniformity monotone.*

Corollary 9 *Consider W , a real-valued function over pairs of probability distributions, and W' , a real-valued function over pairs of quantum states, such that $W'(\rho, \sigma) := W(\lambda(\rho), \lambda(\sigma))$. Then W' is a witness for state conversion under noisy quantum operations if and*

only if W is a witness for state conversion under classical noisy operations.

It may seem surprising that the quantum problem of state conversion reduces to its classical counterpart even though the strict inclusions of the mixtures of unitaries within the noisy operations, and the noisy operations within the unital operations (Lemma 5), have no classical counterparts (Lemma 6). The solution to this puzzle is that, for the purposes of state conversion, the three classes of quantum operations have the same power.

Lemma 10 *Let $\rho \in \mathcal{L}(\mathcal{H}_{\text{in}})$ and $\sigma \in \mathcal{L}(\mathcal{H}_{\text{out}})$. Then, the following propositions are equivalent:*

(i) $\rho \mapsto \sigma$ by a noisy operation

(ii) $\rho \mapsto \sigma$ by a unital operation.

If ρ and σ are of equal dimension, then (i) and (ii) are also equivalent to

(iii) $\rho \mapsto \sigma$ by a mixture of unitaries.

See Appendix A for the proof. For the case of ρ and σ of equal dimension, Uhlmann's Theorem (Uhlmann, 1970) implies that $\rho \mapsto \sigma$ by a mixture of unitaries if and only if the spectrum of ρ majorizes that of σ ; see (Nielsen and Vidal, 2001) for a discussion.¹⁰

There are questions in the resource theory that do not concern state conversion. For instance, one may ask about the possibility of simulating an operation that is outside the free set, given access to some resource state. For such questions, the quantum solution does not necessarily reduce to the classical one. We do not consider such problems in this article. From this point onwards, therefore, we can restrict our attention to the classical resource theory.

III. QUASI-ORDER OF STATES UNDER NOISY OPERATIONS

A. Equivalence classes under noisy operations

The first step in understanding state conversion within the classical resource theory of nonuniformity is to determine when two states are equivalent relative to noisy operations, by which it is meant that they can be reversibly interconverted, one to the other, deterministically, by noisy operations. In this case, we say that the states have precisely the same *nonuniformity properties*.

Definition 11 (Exact state conversion) *We write $x \xrightarrow{\text{noisy}} y$ if there exists a noisy classical operation D such that $y = Dx$.*

⁹ Note, however, that one *does* find a separation between the quantum and the classical theories if one assumes an additional restriction on the free operations, namely, that the systems to which one has access are correlated with others to which one has no access. The reason is that a mixture of entangled states cannot be transformed by a local unitary to a mixture of product states.

¹⁰ The implication from an operation being unital to majorization of the final state's spectrum by the initial state's has also been noted in (Chefles, 2002).

Definition 12 (Noisy-equivalence of states) *Two states, x and y , are said to be noisy-equivalent if they are reversibly interconvertible, that is, if $x \xrightarrow{\text{noisy}} y$ and $y \xrightarrow{\text{noisy}} x$.*

Since noisy classical operations can have an input of dimension d_{in} and an output of dimension d_{out} , state conversion is a map from a vector on $\mathbb{R}_+^{d_{\text{in}}}$ to a vector on $\mathbb{R}_+^{d_{\text{out}}}$. As such, it is useful to introduce the set

$$\mathcal{R} := \bigcup_{d=1}^{\infty} \mathbb{R}_+^d.$$

Any finite-dimensional probability distribution is a vector in \mathbb{R}_+^d for some d , and therefore a member of the set \mathcal{R} . For a distribution $x \in \mathcal{R}$, we will denote by d_x the integer for which $x \in \mathbb{R}_+^{d_x}$.

The simplest case to consider is when x and y are of equal dimension. In this case, if x and y are reversibly interconvertible by noisy operations, there is a permutation that takes one to the other. The proof is as follows. Suppose the doubly stochastic matrix taking x to y is denoted D and the one taking y to x is denoted D' . We can write D (respectively D') as a mixture of permutations π_i (respectively π'_j):

$$D = \sum_i \lambda_i \pi_i, \quad D' = \sum_j \lambda'_j \pi'_j,$$

where $\sum_i \lambda_i = \sum_j \lambda'_j = 1$ and all $\lambda_i, \lambda'_j > 0$. Thus

$$x = D'Dx = \sum_{i,j} \lambda'_j \lambda_i \pi'_j \pi_i x.$$

Consider the convex set consisting of all convex combinations of permutations of x . The state x itself is an extremal point of this convex set. Thus, for all i, j , we must have $\pi'_j \pi_i x = x$, and so $\pi_i x = (\pi'_j)^{-1} x$ is the same state (call it z) for all i . But $y = Dx = \sum_i \lambda_i \pi_i x = z$, so $y = \pi_i x$ is a permutation of x .

If we let x^\downarrow denote the vector having the same components as x but permuted such that they are in descending order, the condition for noisy-equivalence can be expressed as $x^\downarrow = y^\downarrow$.

Note that if x and y are of equal dimension and noisy-equivalent, and if x is the marginal of a correlated state on a larger system, and similarly for y , then because we can get from x to y by a permutation, the conversion can be achieved while preserving all correlations with other systems.

The more interesting case is where x and y have unequal dimension: $d_x \neq d_y$. First note that one can reversibly interconvert x and $x \otimes m$ for any uniform state m on an arbitrary ancilla. We get from x to $x \otimes m$ simply by injecting an ancilla in the uniform state, which is allowed under noisy operations, and we get from $x \otimes m$ to x simply by discarding the ancilla, which is also allowed under noisy operations.

It follows that one can reversibly interconvert x and y by noisy operations if and only if one can reversibly interconvert $x \otimes m$ and $y \otimes m'$, where m and m' are uniform states for any arbitrary pair of ancillas. In particular, this is true if and only if one can reversibly interconvert $x \otimes m$ and $y \otimes m'$ for ancillas having dimensions d_m and $d_{m'}$ such that $d_x d_m = d_y d_{m'}$, so that $x \otimes m$ and $y \otimes m'$ are of equal dimension. As shown above, in this case, the interconversion is possible if and only if there is a permutation that takes $x \otimes m$ to $y \otimes m'$. Given that uniform states are invariant under permutations, this condition is equivalent to

$$x^\downarrow \otimes m = y^\downarrow \otimes m'.$$

This condition, therefore, is necessary and sufficient for reversible interconvertibility of x and y under noisy operations.

In the resource theory of nonuniformity, the only properties of a state that are relevant to determining its value as a resource are its nonuniformity properties, that is, the features of the state that determine its noisy-equivalence class. It is therefore useful to replace the state x by a mathematical object that represents only the nonuniformity properties of x .

Given that x is reversibly interconvertible with $x \otimes m$, this mathematical object must be invariant under adding and removing ancillas in the uniform state. If we plot histograms of x^\downarrow and $(x \otimes m)^\downarrow$, we see that the envelope of each is a step function and that the steps' relative heights are equal. An example is given in Fig. 1. We can make these step functions strictly equal by rescaling them in an appropriate way. The rescaling is chosen such that the area of each bar of the histogram remains the same, while the range of the function becomes $[0, 1]$. Specifically, we define the function $h_x(v)$, with range $v \in [0, 1]$, such that the k th step extends over the range $v \in [(k-1)/d_x, k/d_x]$ and has height $d_x x_k^\downarrow$. Equivalently,

$$h_x(v) \equiv d_x x_{\lfloor d_x v \rfloor}^\downarrow, \quad (3.1)$$

where $\lfloor a \rfloor$ denotes the integer floor of a . Again, an example is provided in Fig. 1. We call $h_x(v)$ the *uniform-rescaled histogram* of x .

Clearly, adding or removing a uniform state m (of arbitrary dimension) leaves the uniform-rescaled histogram invariant, $h_x(v) = h_{x \otimes m}(v), \forall v \in [0, 1]$, as illustrated in Fig. 1. Furthermore, if two states of the same dimension are equal up to a permutation, their uniform-rescaled histograms are equal. In particular, it follows that $x^\downarrow \otimes m = y^\downarrow \otimes m'$ if and only if $h_{x \otimes m}(v) = h_{y \otimes m'}(v) \quad \forall v \in [0, 1]$. But $h_x(v) = h_{x \otimes m}(v)$ so we can conclude that $x^\downarrow \otimes m = y^\downarrow \otimes m'$ holds if and only if $h_x(v) = h_y(v)$. We have shown, therefore, that the uniform-rescaled histogram of a state x is a mathematical object that characterizes the noisy-equivalence class of that state.

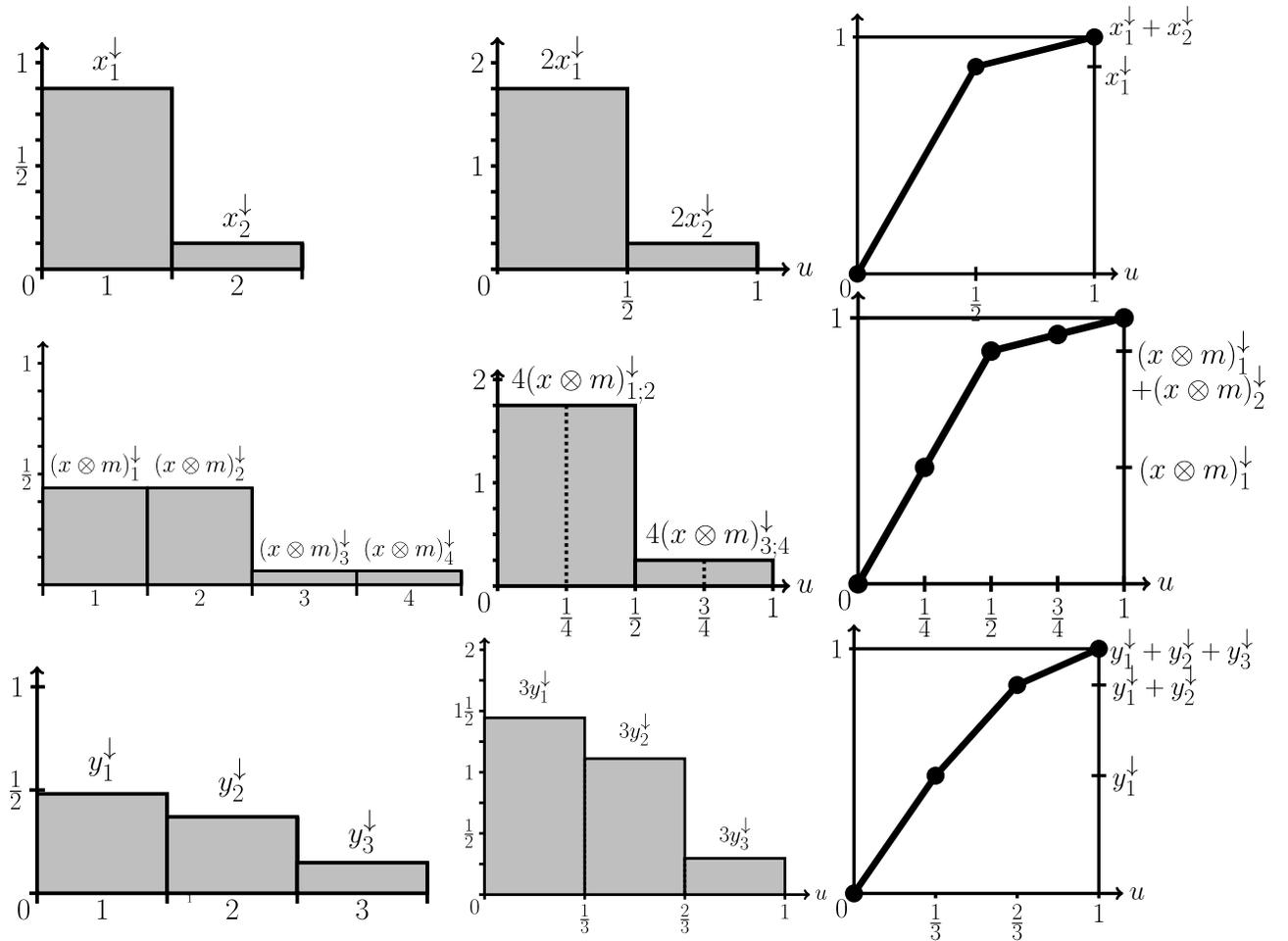


FIG. 1 An illustration of uniform-rescaled histograms and Lorenz curves. The first column depicts the histogram of a state's components, in descending order, the second column depicts the corresponding uniform-rescaled histogram, and the third column depicts the Lorenz curve. The states depicted are: (first row) $x \equiv (0.9, 0.1)$, a state of a 2-level system; (second row) $x \otimes m \equiv (0.45, 0.45, 0.05, 0.05)$, the state of a pair of 2-level systems, where the first is in state x and the second is in the uniform state $m \equiv (0.5, 0.5)$; (third row) y , a generic state of a three-level system.

Now consider the integral of $h_x(v)$ from $v = 0$ to $v = u$ as a function of $u \in [0, 1]$,

$$L_x(u) := \int_0^u h_x(v) dv. \quad (3.2)$$

Clearly, the curve traced by $L_x(u)$ extends between $(0,0)$ and $(1,1)$ regardless of the state x . It contains all the information contained in the rescaled histogram $h_x(v)$. Indeed, one can recover the latter by taking the derivative of $L_x(u)$. As such, it is another mathematical object that characterizes the noisy-equivalence class. Examples of this curve for various states are provided in Fig. 1. For the uniform state m , this curve is simply the diagonal line extending between $(0,0)$ and $(1,1)$.

There is another way of defining this curve which is worth noting. First, define $S_k(x)$ for $k = 1, \dots, d_x$ as the

sum of the k largest components of x ,

$$S_k(x) := \sum_{i=1}^k x_i^\downarrow, \quad (3.3)$$

and define $S_0(x) := 0$ (Note that $S_{d_x}(x) = 1$ for x a normalized probability distribution). $S_k(x)$ is sometimes called the Ky Fan k -norm of x (Bhatia, 1997; Horn and Johnson, 2012). Then we can characterize $L_x(u)$ as the linear interpolation of the points

$$\left(\frac{k}{d_x}, \frac{S_k(x)}{S_{d_x}(x)} \right) \quad \forall k = 0, \dots, d_x. \quad (3.4)$$

$L_x(u)$ is called the *Lorenz curve* of x (Lorenz, 1905), (MOA, Sec. 1.A).¹¹ It will be seen to be one of the

¹¹ Actually, the Lorenz curve of x is conventionally taken to be the linear interpolation of $(k/d_x, T_k/T_{d_x}) \quad \forall k = 0, \dots, d_x$, where

primary tools for characterizing the resource theory of nonuniformity.

It is worth emphasizing that we can also infer that noisy equivalence implies equality of Lorenz curves *directly* from the condition that $x^\downarrow \otimes m = y^\downarrow \otimes m'$. It suffices to note that, via the definition of Lorenz curves in terms of Ky-Fan k -norms, this condition is equivalent to $L_{x \otimes m}(u) = L_{y \otimes m'}(u)$, $\forall u \in [0, 1]$ and then to note that the Lorenz curve is invariant under adding and removing ancillas in a uniform state, $L_{x \otimes m}(u) = L_x(u)$, $\forall u \in [0, 1]$.

To summarize, we have shown that

Proposition 13 (Conditions for noisy equivalence)

A pair of states x and y are noisy-equivalent if and only if the following equivalent conditions holds

(i) *their uniform-rescaled histograms are equal,*

$$h_x(v) = h_y(v), \quad \forall v \in [0, 1], \quad (3.5)$$

(ii) *their Lorenz curves are equal,*

$$L_x(u) = L_y(u) \quad \forall u \in [0, 1]. \quad (3.6)$$

The uniform-rescaled histogram of x and the Lorenz curve of x both capture all and only the nonuniformity properties of x . It follows that for any notion of state conversion we wish to study (exact or approximate; single-shot, multi-copy or asymptotic; catalytic or non-catalytic), the necessary and sufficient conditions under which one state can be converted to another can always be expressed in terms of either of these objects. Furthermore, any nonuniformity monotone or state conversion witness must be expressible entirely in terms of them as well.

The application of these mathematical objects to thermodynamics was first recognized in (Ruch, 1975a,b; Ruch and Lesche, 1978; Ruch and Mead, 1976; Ruch *et al.*, 1978). What we have called the “uniform-rescaled histogram” was discussed in (Ruch *et al.*, 1978) under the title of the “density diagram”. More recently, in (Egloff *et al.*, 2012), these old tools have begun to be used again in the context of an information-theoretic approach to thermodynamics. In this article, an operation of “Gibbs-rescaling” is introduced which is akin to our use of the

uniform-rescaled histogram in place of the distribution itself (although with a different scaling convention for the horizontal axis). The analogue of the Lorenz curve for an athermal state, that is, the generalization of Lorenz curves to a system with a nontrivial Hamiltonian, has recently been studied in (Horodecki and Oppenheim, 2013) and applied in (Brandão *et al.*, 2015).

With this characterization of the noisy equivalence class in hand, we can clarify a point that was made in the introduction, namely, that the nonuniformity properties of a state depend on the dimension of the space in which the state is embedded. As already noted, this is because embedding a state in a higher-dimensional space—that is, padding the state with extra zeros—is *not* a noisy operation. In terms of the uniform-rescaled histogram, padding a state with extra zeros corresponds to squeezing the entire histogram of the state to the left and leaving only zeros on the right side, which obviously results in a different histogram. In the Lorenz curve picture, it corresponds to squeezing the Lorenz curve to the left and adding a horizontal segment at value 1 on the right end, again resulting in something that is obviously distinct from the original Lorenz curve.

B. Deterministic interconversion of nonuniform states

We now turn to a consideration of the necessary and sufficient conditions on a pair of states x and y such that there exists a deterministic noisy operation taking x to y . We are here asking about one-way state conversion, i.e., there need not be any deterministic noisy operation taking y to x . We begin by presenting the general result.

1. The result

Proposition 14 (Conditions for deterministic conversion) $x \xrightarrow{\text{noisy}} y$ *if and only if*

(i) *the uniform-rescaled histogram of x , $h_x(v)$, can be taken to that of y , $h_y(v)$, by moving probability density only towards the right (i.e. from lower to higher values of v).*

(ii) *the Lorenz curve of x is everywhere greater than or equal to the Lorenz curve of y :*

$$L_x(u) \geq L_y(u) \quad \forall u \in [0, 1]. \quad (3.7)$$

When these conditions hold, we say that x noisy-majorizes y .

The rest of this section provides the proof of Proposition 14. We start with the case in which x and y have equal dimension.

Recall the definition of the majorization relation (Definition A.1 of MOA).

$T_k(x) := \sum_{i=1}^k x_i^\uparrow$ is the sum of the k *smallest* components of x (Lorenz, 1905), or equivalently, it is taken to be the integral $L'_x(u) := \int_{1-u}^1 h_x(v) dv$. But the conventional definition is just the inversion about the line extending from (0,0) to (1,1) of our definition, and so the two curves have precisely the same information content. We here choose to adopt the opposite of the usual convention because in this way the majorization relation between states, $x \succ y$, coincides with the inequality relating the height of the Lorenz curves of those states, $L_x(u) \geq L_y(u)$, $\forall u \in [0, 1]$ (see Eq. (3.7) below). Also, our convention coincides with the one adopted in (Horodecki and Oppenheim, 2013).

Definition 15 (Majorization) *Letting x and y be normalized probability vectors with equal dimension d , we say that x majorizes y and write $x \succ y$ if*

$$\sum_{i=1}^k x_i^\downarrow \geq \sum_{i=1}^k y_i^\downarrow \quad \forall k = 1, \dots, d-1. \quad (3.8)$$

Because x and y are normalized probability distributions, $\sum_{i=1}^d x_i^\downarrow = \sum_{i=1}^d y_i^\downarrow = 1$.

The connection with noisy classical operations is made through the following famous result (Hardy *et al.*, 1952):

Lemma 16 (Hardy, Littlewood, Polya) *$x \succ y$ if and only if there is a doubly stochastic matrix D such that $y = Dx$.*

Given that the set of noisy classical operations with equal input and output dimensions are represented by the set of doubly-stochastic matrices (Lemma 6), we immediately obtain that, for x and y of equal dimensions, $x \xrightarrow{\text{noisy}} y$ if and only if $x \succ y$.

An equivalent means of expressing the condition of majorization is in terms of Lorenz curves. Noting that the expressions in the inequalities of Eq. (3.8) are just the Ky Fan norms $S_k(x)$ and recalling that the Lorenz curve is the linear interpolation of points $(k/d_x, S_k(x)/S_{d_x}(x))$, we see that in the case of x and y of equal dimension, we have that $x \succ y$ if and only if the Lorenz curve of x is nowhere less than the Lorenz curve of y :

$$L_x(u) \geq L_y(u) \quad \forall u \in [0, 1]. \quad (3.9)$$

This proves (ii) of Proposition 14 for the case of states of equal dimension.

When x and y are states of unequal dimension, the condition for deterministic conversion is not simply majorization; this is why we use the term *noisy-majorization* to describe the condition in the general case. It is determined using the same trick that was deployed in the characterization of the noisy equivalence classes. It suffices to note that $x \xrightarrow{\text{noisy}} y$ if and only if there exist uniform states m and m' such that $x \otimes m \xrightarrow{\text{noisy}} y \otimes m'$ (because adding and removing uniform states are noisy operations), and that if we choose m and m' such that $d_x d_m = d_y d_{m'}$, then $x \otimes m$ and $y \otimes m'$ are states of equal dimension. If we define d as the least common multiple of d_x and d_y , $d := \text{LCM}(d_x, d_y)$, then it suffices to choose $d_m := d/d_x$ and $d_{m'} := d/d_y$, in which case $x \otimes m$ and $y \otimes m'$ have dimension d .

Recalling that we have established Condition (ii) of Proposition 14 for states of equal dimension, and the fact that the Lorenz curve of $x \otimes m$ for a uniform state m is equal to the Lorenz curve of x , it follows that $x \xrightarrow{\text{noisy}} y$ if and only if the Lorenz curve of x is nowhere below the Lorenz curve of y . This proves Condition (ii) of Proposition 14 for all states.

Finally, recalling that the Lorenz curve is the cumulative integral of the uniform-rescaled histogram, any

motion of density rightward in the uniform-rescaled histogram corresponds to a decrease of the height of the Lorenz curve over some subset of its domain, while motion of density leftward corresponds to an increase of height. Condition (ii) of Proposition 14, therefore, implies (i).

This concludes the proof of Proposition 14.

The application of majorization theory to state conversion in thermodynamics was studied extensively in (Ruch, 1975a,b; Ruch and Lesche, 1978; Ruch and Mead, 1976; Ruch *et al.*, 1978). The quantum information community became familiar with majorization due to its role in the resource theory of entanglement (Nielsen, 1999; Nielsen and Vidal, 2001). The problem of state conversion in thermodynamics was first considered from a quantum information perspective in (Janzing *et al.*, 2000), where some necessary conditions on state conversion were derived. The necessary and sufficient conditions for state conversion under thermal operations were first determined in (Horodecki and Oppenheim, 2013), where the relation was called *thermo-majorization*. The results described in this section are the specialization of the thermo-majorization relation to the case of a trivial Hamiltonian.

2. Some consequences

The order over states induced by deterministic conversion is not a total order but a quasi-order. We call it the *noisy quasi-order*. One easily generates pairs of states that are not noisy-ordered relative to one another by simply drawing a pair of valid Lorenz curves where one is not everywhere above the other.¹²

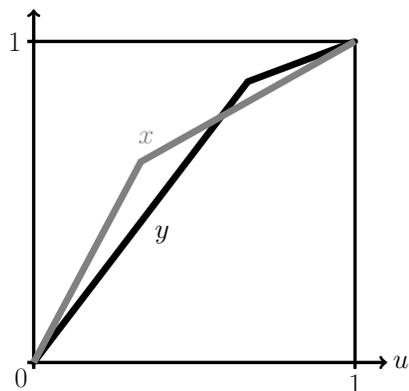


FIG. 2 The Lorenz curves of a pair of states, x and y , such that it is neither the case that $x \xrightarrow{\text{noisy}} y$ nor that $y \xrightarrow{\text{noisy}} x$.

¹² The order is a *quasi-order* (also known as a pre-order) rather than a partial order, because we can have $x \succ y$ and $y \succ x$ for $x \neq y$. This occurs whenever x is a nontrivial permutation of y or requires addition or removal of a uniform state. While the states form a quasi-order, the noisy-equivalence classes of states form a partial order.

Note that one can easily recover the condition for noisy-equivalence (Proposition 13) from the condition for noisy-majorization (Proposition 14) by recognizing that reversible interconvertibility requires noisy-majorization in both directions.

Another simple corollary of the deterministic-conversion result concerns the relation between a state of a composite and its marginal state. Suppose x^{AB} is a state of a composite system $\Omega_A \times \Omega_B$ and x^A is the marginal state on Ω_A , that is, $x_i^A = \sum_j x_{ij}^{AB}$, where $i \in \Omega_A, j \in \Omega_B$. x^A is noisy-majorized by x^{AB} , that is, $L_{x^{AB}}(u) \geq L_{x^A}(u)$ for all $u \in [0, 1]$.

For noisy-equivalence of x^{AB} and x^A , we require that $L_{x^{AB}}(u) = L_{x^A}(u)$ which implies that $x^{AB\downarrow} = x^{A\downarrow} \otimes m^B$. It follows that marginalization is reversible only if the marginalized system is uncorrelated with the rest and is in a uniform state.

Proposition 14 also implies that the height of the Lorenz curve at a given value of u in the region $[0, 1]$ is a nonuniformity monotone, $M_{L_{\text{Lorenz}, u}}(x) \equiv L_x(u)$, and the set of such heights, $\{M_{L_{\text{Lorenz}, u}} : u \in [0, 1]\}$, form a complete set of nonuniformity monotones. Although this is an infinite set, if x and y are both of finite dimension (the only case we consider in this article), one can decide the convertibility question by looking at a finite number of monotones. The following is the pertinent result.

Lemma 17 *For x and y of finite dimension, x noisily-majorizes y if and only if $L_x(u) \geq L_y(u)$ at the points $u = k/d_y$ for all $k = 1, \dots, d_y - 1$. In other words, it suffices to consider the $d_y - 1$ monotones*

$$M_{L_{\text{Lorenz}, k/d_y}}(x) := L_x(k/d_y), \quad k = 1, \dots, d_y - 1. \quad (3.10)$$

The proof is simply that $L_y(u)$ is linear between the distinguished values of u . Given that $L_x(u)$ is concave, if $L_x(u) \geq L_y(u)$ at these points, then $L_x(u) \geq L_y(u)$ at all $u \in [0, 1]$.

This result can be rephrased in terms of state-conversion witnesses as follows.

Corollary 18 *Each of the functions*

$$\Delta_k(x||y) := L_x(k/d_y) - L_y(k/d_y) \quad (3.11)$$

for $k \in \{1, \dots, d_y\}$ is a no-go witness for $x \xrightarrow{\text{noisy}} y$. That is, if $\Delta_k(x||y) < 0$ then it is not the case that $x \xrightarrow{\text{noisy}} y$.

The function

$$\Delta(x||y) := \min_{k \in \{1, \dots, d_y\}} \Delta_k(x||y) \quad (3.12)$$

is a complete witness for the state conversion. That is, $x \xrightarrow{\text{noisy}} y$ if and only if $\Delta(x||y) \geq 0$.

In Section V.E, we will discuss how this result implies the inadequacy of the standard formulation of the second law of thermodynamics.

3. Implementation

In the previous section, we determined the conditions under which it is possible to transform x to y by a noisy operation, but the proof was not constructive. In this section, we describe a practical implementation of the appropriate noisy operation. We begin with states of equal dimensions.

First of all, we recall the definition of majorization in terms of a sequence of *T-transforms*. A T-transform is a doubly-stochastic matrix that is nontrivial on a single 2×2 block. For the block corresponding to levels i and j , we denote the T-transform by T_{ij} . The only permutations that act only on levels i and j are the identity, denoted I , and the permutation that swaps i and j , which we denote by Π_{ij} . Therefore, by Birkhoff's Theorem (Birkhoff, 1946), the most general form of T_{ij} is

$$T_{ij} = wI + (1 - w)\Pi_{ij},$$

where $0 \leq w \leq 1$. It follows that if x and y are d -dimensional vectors and $y = T_{ij}x$, then

$$y_i = wx_i + (1 - w)x_j, \quad (3.13)$$

$$y_j = wx_j + (1 - w)x_i. \quad (3.14)$$

Equivalently, defining $\bar{x}_{ij} := \frac{1}{2}(x_i + x_j)$ and $q := 2|w - \frac{1}{2}|$, we have

$$y_i = qx_i + (1 - q)\bar{x}_{ij}, \quad (3.15)$$

$$y_j = qx_j + (1 - q)\bar{x}_{ij}. \quad (3.16)$$

We see that if $w = 1/2$ ($q = 0$), then the weights of the pair of levels i and j become equal, and if $w \neq 0, 1$ ($q \neq 1$), then these weights become closer to equal (assuming they were unequal to begin with). In the context of income inequality, a T-transform is called a ‘‘Robin Hood transfer’’ (Arnold, 1987).

Lemma 19 (Muirhead, Hardy, Littlewood, Polya)

For states of equal dimensions, $x \succ y$ if and only if there is a finite sequence of T-transforms taking x to y . The number of steps required is at most $d - 1$, where d is the dimension.

Proof. We follow the proof in (MOA, Lemma B.1, p. 32). Let $x^{(n)}$ denote the state after the n th step of the sequence of T-transforms, so that $x^{(0)} = x$ and $x^{(n_{\text{max}})} = y$. Consider the step that takes $x^{(n)}$ to $x^{(n+1)}$. Let j_{ex} denote the largest index such that the weight for that index is strictly larger for $x^{(n)\downarrow}$ than it is for $x^{(n+1)\downarrow}$, that is, $x_{j_{\text{ex}}}^{(n)\downarrow} > x_{j_{\text{ex}}}^{(n+1)\downarrow}$ (‘‘ex’’ denotes ‘‘excess’’). Let j_{df} denote the smallest index such that the weight for that index is strictly smaller for $x^{(n)\downarrow}$ than it is for $x^{(n+1)\downarrow}$, that is, $x_{j_{\text{df}}}^{(n)\downarrow} < x_{j_{\text{df}}}^{(n+1)\downarrow}$ (‘‘df’’ denotes ‘‘deficient’’). By definition, we must have $j_{\text{ex}} < j_{\text{df}}$. This is illustrated in Fig. 3.

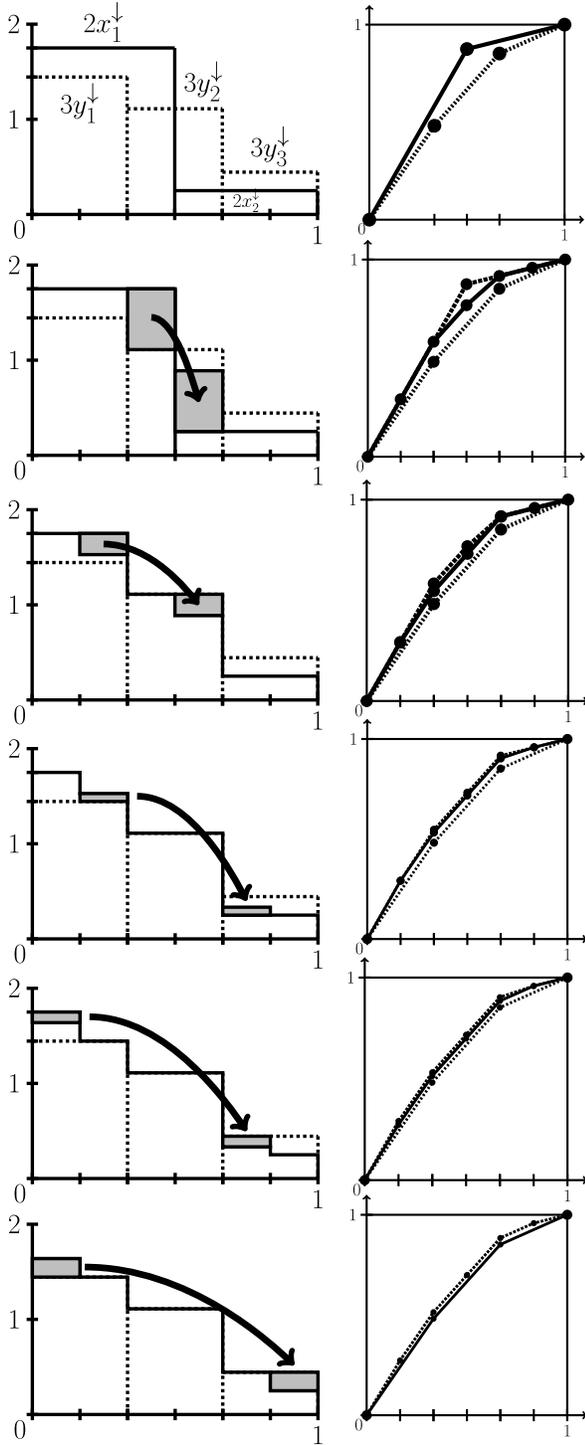


FIG. 3 A depiction of the evolution of the state during a minimal sequence of T-transforms achieving the state conversion $x \xrightarrow{\text{noisy}} y$. The left column denotes the uniform-rescaled histograms of the states, while the right depicts their Lorenz curves.

We consider a protocol where, in the n^{th} step in the sequence, one transfers the maximum weight possible from level j_{ex} to level j_{df} while ensuring that one still has

$x_j^{(n)\downarrow} \geq x_j^{(n+1)\downarrow}$ for $j = j_{\text{ex}}$ and for $j = j_{\text{df}}$ (so that, in particular, $x^{(n)}$ still majorizes $x^{(n+1)}$). This maximum weight, denoted δ , cannot exceed the difference of weights in level j_{ex} nor the difference of weights in level j_{df} , so

$$\delta = \min\{x_{j_{\text{ex}}}^{(n)\downarrow} - x_{j_{\text{ex}}}^{(n+1)\downarrow}, x_{j_{\text{df}}}^{(n+1)\downarrow} - x_{j_{\text{df}}}^{(n)\downarrow}\}.$$

The resulting transformation is

$$x_{j_{\text{ex}}}^{(n+1)\downarrow} = x_{j_{\text{ex}}}^{(n)\downarrow} - \delta, \quad (3.17)$$

$$x_{j_{\text{df}}}^{(n+1)\downarrow} = x_{j_{\text{df}}}^{(n)\downarrow} + \delta. \quad (3.18)$$

Because the transfer tends to make the weights of the two levels closer to equal, it is clearly a T-transform. Specifically, it is the transform $T_{j_{\text{ex}}j_{\text{df}}} = wI + (1-w)\Pi_{j_{\text{ex}}j_{\text{df}}}$, where $w = 1 - \delta/(x_{j_{\text{ex}}}^{(n)\downarrow} - x_{j_{\text{df}}}^{(n)\downarrow})$. After the n^{th} step of the sequence, either the index j_{ex} is reduced by 1 or the index j_{df} is increased by 1 (or both), so that at the next step, one is transferring weight between a different pair of levels. Clearly, if x majorizes y , then a sequence of such transformations can take x to y in a finite number of steps.

Finally, we show that at most $d-1$ steps are required. Suppose $d(x, y)$ denotes the number of levels wherein x^\downarrow and y^\downarrow differ in value. Because the number of differences is reduced by 1 at every step of the sequence, and the last step takes two differences to no differences, it follows that one requires $d(x, y) - 1$ steps. $d(x, y)$, however, is at most d . ■

This covers the case of states of equal dimension. Next, we consider how to implement a noisy operation that achieves deterministic conversion of states of unequal dimensions. It follows from Lemma 19 that we can achieve the transformation by a sequence of T-transforms, with the number of steps in the sequence being at most $d-1$ where $d = \text{LCM}(d_x, d_y)$. Fig. 3 depicts the set of T-transforms that maps $x \otimes m_1$ to $y \otimes m_2$ (where $d_{m_1} := d/d_x$ and $d_{m_2} := d/d_y$) by the protocol described in the proof of Lemma 19. Note that we must divide the x -axis into bins of size $1/d$ to depict how the sequence of T-transforms acts on the uniform-rescaled histograms. Note also that this figure provides an intuitive proof of Proposition 14 (i). Finally, note how the T-transforms act on the Lorenz curves. Each T-transform acts on the pair of bins for which the difference between the slope differential of the Lorenz curves in the first bin and the slope differential of the Lorenz curves in the second bin is largest. After a particular T-transform is complete, the slope differential in one of the two bins becomes zero, and consequently, the Lorenz curves have the same slope in that bin. By this process, the Lorenz curve of the initial state approaches that of the target state.

IV. NONUNIFORMITY MONOTONES

In this section, we discuss the properties and method of construction of functions which serve as monotones

under classical noisy operations. A reader interested in getting a quick introduction to nonuniformity monotones (without too much technical detail) will find the tables in this section useful: Table I lists some monotones derived from convex functions of one real variable, outlining the steps of the derivation; Table II lists some monotones which can be derived from the geometry of the Lorenz curve; and Table III lists the operational significance of some of the monotones.

Given that noisy operations may change the dimension of the space, every nonuniformity monotone is really a family of functions, $\{G_d : d \in \mathbb{Z}_+\}$, where $G_d : \mathbb{R}_+^d \rightarrow \mathbb{R}$ is defined on a d -dimensional space. We will denote by $G : \mathcal{R} \rightarrow \mathbb{R}$ the function which reduces to G_d on \mathbb{R}_+^d . Thus, $\forall x \in \mathcal{R}$,

$$G(x) = G_{d_x}(x). \quad (4.1)$$

Where the dimensionality is clear from the context, we will omit the subscript x in d_x .

A monotone is said to be *strict* if $M(x) = M(y)$ only on level sets, that is, only when x and y are either noisy-equivalent (i.e., $x \xrightarrow{\text{noisy}} y$ and $y \xrightarrow{\text{noisy}} x$) or not ordered relative to one another (i.e., it is neither the case that $x \xrightarrow{\text{noisy}} y$ nor that $y \xrightarrow{\text{noisy}} x$).

Since appending ancillary systems in the uniform state is a reversible process under noisy operations, every nonuniformity monotone M must satisfy $M(x \otimes m^{(d)}) = M(x) \forall d \in \mathbb{Z}_+$, where $m^{(d)}$ is the uniform state of dimension d .

Within a space of fixed dimension, the noisy quasi-order reduces to the majorization quasi-order, so any nonuniformity monotone M , when restricted to states of a fixed dimension, must be a nonincreasing monotone with respect to the majorization quasi-order. The latter are known as *Schur-convex* functions and have been extensively studied. See, e.g., Chapter 3 of MOA.

Definition 20 (Schur-convexity) A function G_d mapping d -dimensional probability distributions to the reals is Schur-convex iff for all d -dimensional distributions x and y ,

$$x \succ y \Rightarrow G_d(x) \geq G_d(y). \quad (4.2)$$

As it turns out, for a function $M : \mathcal{R} \rightarrow \mathbb{R}$, the conditions that the restriction of M to every d -dimensional vector space be Schur-convex and that M be invariant under the addition of a uniform state are necessary and sufficient for M to be a nonuniformity monotone. We formalize this in the following.

Proposition 21 A function $M : \mathcal{R} \rightarrow \mathbb{R}$ is a nonuniformity monotone if and only if both of the following conditions hold:

1. For each $d \in \mathbb{Z}_+$, the restriction of M to \mathbb{R}_+^d is Schur-convex.

2. For all $d \in \mathbb{Z}_+$, for all distributions $x \in \mathcal{R}$,

$$M(x \otimes m^{(d)}) = M(x), \quad (4.3)$$

where $m^{(d)}$ is the d -dimensional uniform distribution.

Proof. The forward implication is trivial, so our task is to show that the pair of conditions imply that M is a nonuniformity monotone.

We begin by noting that for each d , if the restriction of M to \mathbb{R}_+^d is Schur-convex, then it follows by definition that if x and y are states of equal dimension, and x majorizes y , then $M(x) \geq M(y)$. It remains to show that if x and y are of unequal dimension, and x noisy-majorizes y , then $M(x) \geq M(y)$.

By assumption, $M(x \otimes m^{(d)}) = M(x)$, for any d . Then, one can reason as follows. First, we can define states $x \otimes m^{(d_y)}$ and $y \otimes m^{(d_x)}$ that are of equal dimension, namely $d_x d_y$. Given that x noisy-majorizes y , and given that adding uniform states does not change the noisy-equivalence class of a state, it follows that $x \otimes m^{(d_y)}$ noisy-majorizes $y \otimes m^{(d_x)}$. But since noisy-majorization between states of equal dimension is just majorization, it follows that $x \otimes m^{(d_y)}$ majorizes $y \otimes m^{(d_x)}$. Then, by the Schur-convexity of the restriction of M to $\mathbb{R}^{d_x d_y}$, we conclude that $M(x \otimes m^{(d_y)}) \geq M(y \otimes m^{(d_x)})$. It follows that $M(x) \geq M(y)$. ■

Note that if one takes an *arbitrary* family of Schur-convex functions, $\{G_d : d \in \mathbb{Z}_+\}$ where $G_d : \mathbb{R}_+^d \rightarrow \mathbb{R}$, then the function $G : \mathcal{R} \rightarrow \mathbb{R}$ which reduces to G_d on \mathbb{R}_+^d need not be a nonuniformity monotone because it need not satisfy the requirement of invariance under adjoining a uniform state, Eq. (4.3). So the problem of defining nonuniformity monotones is not solved by merely finding families of Schur-convex functions. Nonetheless, the powerful characterization theorems for Schur-convex functions can be exploited to construct and characterize nonuniformity monotones. We will discuss these theorems in the following.

A. Nonuniformity monotones built from convex functions on the reals

1. Schur-convexity

The following result is well-known; cf. (MOA, Proposition C.1).

Lemma 22 For every convex function $g : \mathbb{R}_+ \rightarrow \mathbb{R}$, the function $G_d : \mathbb{R}_+^d \rightarrow \mathbb{R}$ defined by

$$G_d(x) := \sum_{i=1}^d g(x_i)$$

is Schur-convex for each $d \in \mathbb{Z}_+$. That is, if x and y are of equal dimension d , and if $x \succ y$, then $G_d(x) \geq G_d(y)$.

Proof. Recall that a function $g : \mathbb{R} \rightarrow \mathbb{R}$ is convex if, for any pair of points $a, a' \in \mathbb{R}$,

$$g(wa + (1-w)a') \leq wg(a) + (1-w)g(a')$$

for all $w \in [0, 1]$. If g is twice differentiable, an equivalent definition of convexity is that $g''(a) \geq 0$ for all $a \in \mathbb{R}$.

Note also that the function $G_d(x)$ is invariant under permutation of the components of x . Hence it is sometimes described as a symmetric function.

The proof is then straightforward. Given that $x \succ y$, it follows that $y = Dx$ for some doubly-stochastic matrix D . By Birkhoff's Theorem (Birkhoff, 1946), there exists a probability distribution (w_i) and a set of permutations (Π_i) such that $y = \sum_i w_i \Pi_i x$. Because G_d is a sum of convex functions, it is also convex, and therefore, $G_d(y) \leq \sum_i w_i G_d(\Pi_i x)$. But we noted above that G_d is invariant under a reordering of the components of its argument, so that $G_d(\Pi_i x) = G_d(x)$. It follows that $G_d(y) \leq \sum_i w_i G_d(x) = G_d(x)$. ■

Now that we have seen how to generate families of Schur-convex functions, the question arises of which of these families can yield a function that is invariant under adjoining uniform states.

2. Schur-convexity relative to a distribution q

Towards this end, we introduce the notion of majorization relative to a distribution q ; see (MOA, Ch. 14, Sec. B, p. 585).

Within a space of a given dimension d , consider those stochastic matrices that preserve a particular distribution q . Call these the q -preserving stochastic matrices.

Definition 23 (Majorization relative to q) A distribution x is said to majorize another distribution y relative to the distribution q , denoted $x \succ_q y$, if there exists a q -preserving stochastic matrix D such that $y = Dx$.

Definition 24 (Schur-convexity relative to q) We say that a function $f : \mathbb{R}_+^d \rightarrow \mathbb{R}$ is Schur-convex relative to q if $x \succ_q y$ implies $f(x) \geq f(y)$.

Intuitively, whereas a Schur-convex function quantifies the distance of some distribution x to the uniform distribution m , a Schur-convex function relative to q quantifies the distance of x to q .

The following is a useful method of constructing Schur-convex functions relative to a distribution q , from convex functions of one real variable.

Lemma 25 Given a distribution $q \in \mathbb{R}_+^d$ with all $q_i \neq 0$, for every convex function $g : \mathbb{R}_+ \rightarrow \mathbb{R}$, the function $G(\cdot \| q) : \mathbb{R}_+^d \rightarrow \mathbb{R}$ defined by

$$G(x \| q) := \sum_{i=1}^d q_i g\left(\frac{x_i}{q_i}\right) \quad (4.4)$$

is Schur-convex relative to q . That is, if x and y are of equal dimension d , and if $x \succ_q y$, then $G(x \| q) \geq G(y \| q)$.

This is Proposition B.3 in Ch. 3 of MOA, proven in (Veinott, 1971). Functions of the form of Eq. (4.4) have also been proposed as a generalization of the notion of relative entropy in (Csiszár, 1967) (where it is called the g -divergence of x from q) and in (Morimoto, 1963), as discussed in (Gorban *et al.*, 2010).

Note that if we multiplied each q_i by a factor $c > 0$ in Eq. (4.4), we would still have a Schur-convex function relative to q (because the function $g'(a) := cg(a/c)$ is also convex). Nonetheless, the case of $c = 1$ has a special status. Suppose we define a fiducial distribution q for every system and suppose that the fiducial distribution on a composite system is the product of the fiducial distributions on the components, that is, $q^{AB} = q^A \otimes q^B$. In this case, we can prove the following.

Lemma 26 For every convex function $g : \mathbb{R}_+ \rightarrow \mathbb{R}$, the function $M_g : \mathcal{R} \rightarrow \mathbb{R}$ defined by $M_g(x) \equiv G(x \| q)$ satisfies $M_g(x \otimes q^S) = M_g(x)$ for any system S .

Proof. It suffices to note that

$$G(x^A \otimes q^B \| q^{AB}) = G(x^A \otimes q^B \| q^A \otimes q^B) \quad (4.5)$$

$$= \sum_{i,j} q_i^A q_j^B g\left(\frac{x_i^A q_j^B}{q_i^A q_j^B}\right) \quad (4.6)$$

$$= \sum_i q_i^A g\left(\frac{x_i^A}{q_i^A}\right) \quad (4.7)$$

$$= G(x^A \| q^A), \quad (4.8)$$

where we have used the fact that $\sum_j q_j^B = 1$. ■

It follows that any family of functions (one for every type of system), each of which is Schur-convex relative to the fiducial distribution on that system, can be used to construct resource monotones in a classical resource theory where those fiducial distributions are the free states.

In particular, in the context of the resource theory of athermality (Brandão *et al.*, 2015, 2013), where the free states are thermal states, we can construct athermality monotones from families of functions that are Schur-convex relative to the thermal state (note that the thermal state of a system depends not only on the dimension of that system, but also on its Hamiltonian).

3. Schur-convexity relative to the uniform state

By specializing to the case where the fiducial distribution for a system of dimension d is just the uniform distribution of that dimension, Lemma 26 provides a method of constructing a nonuniformity monotone from any convex function.

Theorem 27 For every convex function $g : \mathbb{R}_+ \rightarrow \mathbb{R}$, the function $M_g : \mathcal{R} \rightarrow \mathbb{R}$ defined by

$$M_g(x) := G\left(x \parallel m^{(d_x)}\right) = \frac{1}{d_x} \sum_{i=1}^{d_x} g(d_x x_i)$$

is a nonuniformity monotone.

Proof. We must show that the two conditions of Proposition 21 hold. It is given that $M_g(x) = G\left(x \parallel m^{(d_x)}\right)$, therefore it follows from Theorem 26 that for any $d' \in \mathbb{Z}_+$, $M_g\left(x \otimes m^{(d')}\right) = M_g(x)$, and it follows from Theorem 25 that for any $d \in \mathbb{Z}_+$, the restriction of M_g to a d -dimensional space is Schur-convex relative to $m^{(d)}$, hence Schur-convex. ■

Note that if a Schur-convex function G can be expressed in the form $G(x) := \sum_{i=1}^{d_x} g(x_i)$ for some function $g : \mathbb{R} \rightarrow \mathbb{R}$, then g is convex [Proposition C.1.c on p. 95 of MOA]. Therefore, the theorem above defines a nonuniformity monotone for every Schur-convex function that is expressible in this form.

Recall that a set of nonuniformity monotones that determines the noisy equivalence class of a state is called a *complete* set of monotones. The set of monotones M_g that is defined by the set of *all* convex functions g is a complete set (Hardy *et al.*, 1952). We have also seen that the monotones describing the height of the Lorenz curve at every point are a complete set. For other examples, see Ch. 4, Sec. B of MOA, and (Joe, 1990; Ruch *et al.*, 1978).

4. Schur-convexity-preserving functions

It is easy to see directly from Definition 20 that the composition of any Schur-convex function $G_d : \mathbb{R}_+^d \rightarrow \mathbb{R}$ with a function $f : \mathbb{R} \rightarrow \mathbb{R}$ that is *increasing* on its domain gives another Schur-convex function $f \circ G_d : \mathbb{R}_+^d \rightarrow \mathbb{R}$. Similarly, the composition of any Schur-convex function relative to q , $G(\cdot \parallel q) : \mathbb{R}_+^{d_q} \rightarrow \mathbb{R}$, with an increasing function f gives another function that is Schur-convex relative to q , $f \circ G(\cdot \parallel q) : \mathbb{R}_+^{d_q} \rightarrow \mathbb{R}$.

Schur-convex functions of x can be understood as generalizations of the negative entropy, or “negentropy”, of x and are therefore denoted by $-H_{(\cdot)}(x)$, with a subscript labelling the particular such function. Similarly, functions of x that are Schur-convex relative to q can be naturally understood as generalizations of the relative entropy of x to q and are therefore denoted by $H_{(\cdot)}(x \parallel q)$. Finally, nonuniformity monotones of x will be denoted $I_{(\cdot)}(x)$.

5. Examples, including order- p Rényi nonuniformities

Table I provides examples of nonuniformity monotones, constructed from various convex functions using Theorem 27. These are derived as monotones for the classical resource theory, but, by Corollary 8, we can evaluate these monotones for the spectrum of a quantum state to obtain a monotone for the quantum resource theory.

Suppose we take our convex function $g : \mathbb{R}_+ \rightarrow \mathbb{R}$ to be

$$g(a) := a^2.$$

We first construct from this function a family of Schur-convex functions $\{G_d : d \in \mathbb{Z}_+\}$ where $G_d : \mathbb{R}_+^d \rightarrow \mathbb{R}$, using the method of Lemma 22:

$$G_d(x) := \sum_{i=1}^d x_i^2.$$

We then derive, for each G_d , the associated Schur-convex function relative to some distribution $q \in \mathcal{R}$, denoted $G(\cdot \parallel q)$:

$$G(x \parallel q) := \sum_{i=1}^{d_q} \frac{x_i^2}{q_i}.$$

Finally, we construct a nonuniformity monotone $M_g : \mathcal{R} \rightarrow \mathbb{R}$ from this function by taking q to be the uniform distribution of dimension equal to that of the argument:

$$M_g(x) := G(x \parallel m^{(d_x)}) = d_x \sum_{i=1}^{d_x} x_i^2.$$

The reader can easily follow this process of definition for each of the examples provided in Table I. We here provide some comments on these.

The Amato index is the nonuniformity monotone defined from the convex function $g(a) := \sqrt{1 + a^2}$. It is a simple example of a monotone which is also easily justified by considering the geometry of the Lorenz curve, as we will demonstrate in Section IV.B.

Perhaps the most paradigmatic example of a Schur-convex function is the negative of the Shannon entropy, equivalently, the Shannon *negentropy*, which is generated by Lemma 22 from the convex function $g(a) = a \log a$ (in this article, all logarithms are base 2). Using the standard notational convention of $H(x)$ for the Shannon entropy, the Shannon negentropy is simply

$$-H(x) := \sum_{i=1}^{d_x} x_i \log x_i.$$

The relative entropy that $g(a) = a \log a$ defines is

$$H(x \parallel q) := \sum_{i=1}^{d_x} x_i \log \left(\frac{x_i}{q_i} \right),$$

A sampling of nonuniformity monotones derived from convex functions on the reals

Function pair	Schur-convex function on \mathbb{R}_+^d	Schur-convex function relative to q	Classical nonuniformity monotone
$g(a)$ (convex); $f(t)$ (increasing)	$-H_{(\cdot)}(x) := f \circ G_d(x)$ $= f \left(\sum_{i=1}^d g(x_i) \right)$	$H_{(\cdot)}(x q) := f \circ G(x q)$ $= f \left(\sum_{i=1}^{d_q} q_i g \left(\frac{x_i}{q_i} \right) \right)$	$I_{(\cdot)}(x) := H_{(\cdot)} \left(x \parallel m^{(d_x)} \right)$ $= f \left(\sum_{i=1}^{d_x} \frac{1}{d_x} g(d_x x_i) \right)$
$g(a) = a^2$ $f(t) = t$	$\sum_{i=1}^d x_i^2$	$\sum_{i=1}^{d_q} \frac{x_i^2}{q_i}$	$d_x \sum_{i=1}^{d_x} x_i^2$
$\sqrt{1+a^2}$ t	$\sum_{i=1}^d \sqrt{1+x_i^2}$	$\sum_{i=1}^{d_q} q_i \sqrt{1+\frac{x_i^2}{q_i^2}}$	Amato index $I_{\text{Amato}}(x) := \sum_{i=1}^{d_x} \sqrt{\frac{1}{d_x^2} + x_i^2}$
$a \log a$ t	Shannon negentropy $-H(x) := \sum_{i=1}^d x_i \log x_i$	Relative Shannon entropy to q $H(x q) := \sum_{i=1}^{d_q} x_i \log \left(\frac{x_i}{q_i} \right)$	Shannon nonuniformity $I(x) := H \left(x \parallel m^{(d_x)} \right)$ $= \log d_x - H(x)$
$-\log a$ t	Burg negentropy $-H_{\text{Burg}}(x) := -\sum_{i=1}^d \log x_i$	Relative Burg entropy to q $H_{\text{Burg}}(x q) := \sum_{i=1}^{d_q} q_i \log \left(\frac{q_i}{x_i} \right)$ $(= H(q x))$	Burg nonuniformity $I_{\text{Burg}}(x) := H_{\text{Burg}} \left(x \parallel m^{(d_x)} \right)$ $= \log d_x - H_{\text{Burg}}(x)$ $(= H \left(m^{(d_x)} \parallel x \right))$
$\pm a^p$ $\frac{\text{Sgn}(p)}{p-1} (\pm t - 1)$ $p \in \mathbb{R} \setminus \{0,1\}$	Order- p Tsallis negentropy $-H_p^{\text{Ts}}(x) := \frac{\text{Sgn}(p)}{p-1} \left(\sum_{i=1}^d x_i^p - 1 \right)$	Order- p Relative Tsallis entropy to q $H_p^{\text{Ts}}(x q) := \frac{\text{Sgn}(p)}{p-1} \left(\sum_{i=1}^{d_q} x_i^p q_i^{1-p} - 1 \right)$	Order- p Tsallis nonuniformity $I_p^{\text{Ts}}(x) := H_p^{\text{Ts}} \left(x \parallel m^{(d_x)} \right)$ $= d_x^{p-1} \left(H_p^{\text{Ts}} \left(m^{(d_x)} \right) - H_p^{\text{Ts}}(x) \right)$
$\pm a^p$ $\frac{\text{Sgn}(p)}{p-1} \log(\pm t)$ $p \in \mathbb{R} \setminus \{0,1\}$	Order- p Rényi negentropy $-H_p(x) := \frac{\text{Sgn}(p)}{p-1} \log \sum_{i=1}^d x_i^p$	Order- p Relative Rényi entropy to q $H_p(x q) := \frac{\text{Sgn}(p)}{p-1} \log \left(\sum_{i=1}^{d_q} x_i^p q_i^{1-p} \right)$	Order- p Rényi nonuniformity $I_p(x) := H_p \left(x \parallel m^{(d_x)} \right)$ $= \text{Sgn}(p) \log d_x - H_p(x)$
$\delta_{a,0} - 1 \equiv -a^0$ $-\log(-t)$	Order-0 Rényi negentropy (neg-max-entropy) $-H_0(x) := -\log \text{supp}(x) $	Order-0 Relative Rényi entropy (relative max-entropy) to q $H_0(x q) := -\log \left(\sum_{i \in \text{supp}(x)} q_i \right)$	Order-0 Rényi nonuniformity $I_0(x) := H_0 \left(x \parallel m^{(d_x)} \right)$ $= \log d_x - H_0(x)$
	Order- ∞ Rényi negentropy (neg-min-entropy) $-H_\infty(x) := \log x_1^\downarrow$	Order- ∞ Relative Rényi entropy (relative min-entropy) to q $H_\infty(x q) := \log \max_{i=1, \dots, d_q} \left(\frac{x_i}{q_i} \right)$	Order- ∞ Rényi nonuniformity $I_\infty(x) := H_\infty \left(x \parallel m^{(d_x)} \right)$ $= \log d_x - H_\infty(x)$
	Order- $(-\infty)$ Rényi negentropy $-H_{-\infty}(x) := -\log \min_{i=1, \dots, d} x_i$ $= -\log x_d^\downarrow$	Order- $(-\infty)$ Relative Rényi entropy $H_{-\infty}(x q) := -\log \min_{i=1, \dots, d_q} \left(\frac{x_i}{q_i} \right)$ $(= H_\infty(q x))$	Order- $(-\infty)$ Rényi nonuniformity $I_{-\infty}(x) := H_{-\infty} \left(x \parallel m^{(d_x)} \right)$ $= -\log d_x - H_{-\infty}(x)$ $(= H_\infty \left(m^{(d_x)} \parallel x \right))$

TABLE I The first column specifies a function g which is convex and another function f which is increasing (on the relevant domains of definition), and the second specifies the Schur-convex family that they define. In the third column, we define the associated relative Schur-convex function to a distribution q . Finally, in the fourth column, we define a nonuniformity monotone as the relative Schur-convex function to the uniform state. The last four rows concern the nonuniformity monotones based on Rényi entropies. The symbol \pm denotes a sign that depends on p . It is $+$ for $p < 0$ and $p > 1$, and $-$ for $0 < p < 1$ (formally, it equals $\text{Sgn}(p(p-1))$).

which we term the *relative Shannon entropy* of x to q . It is also known as the Kullback–Leibler divergence of x from q (Kullback and Leibler, 1951). This yields the nonuniformity monotone

$$\begin{aligned} I(x) &:= H(x||m) = \sum_{i=1}^{d_x} x_i \log(d_x x_i) \\ &= \log d_x - H(x), \end{aligned}$$

which we term the *Shannon nonuniformity* of x .

Starting from the convex function $g(a) = -\log a$, we obtain as the corresponding Schur-convex function the negative of the *Burg entropy* (Burg, 1967),

$$-H_{\text{Burg}}(x) := -\sum_{i=1}^{d_x} \log x_i.$$

We find the relative Schur-convex function that this defines to be

$$H_{\text{Burg}}(x||q) := \sum_{i=1}^{d_x} q_i \log\left(\frac{q_i}{x_i}\right).$$

It is straightforward to verify that

$$H_{\text{Burg}}(x||q) = H(q||x).$$

So the relative Burg entropy of x to q is just the relative Shannon entropy of q to x .

In the context of athermality theory, where q is the thermal state, the analogous quantity has been discussed in (Janzing *et al.*, 2000).

As the next example, consider the one-parameter family of functions

$$g_p(a) := \pm a^p \quad (p \in \mathbb{R} \setminus \{0, 1\}), \quad (4.9)$$

where the sign is chosen such that g_p is convex (that is, $+$ for $p < 0$ and $p > 1$, and $-$ for $0 < p < 1$). We first obtain that $G^{(p)}(x) := \pm \sum_{i=1}^{d_x} x_i^p$ is Schur-convex. Setting $f(t) := \text{Sgn}(p)(\pm t - 1)/(p - 1)$ yields the *order- p Tsallis negentropy* (Tsallis, 1988) $-H_p^{\text{Ts}} := f \circ G^{(p)}$; that is,

$$-H_p^{\text{Ts}}(x) := \frac{\text{Sgn}(p)}{p - 1} \left(\sum_{i=1}^{d_x} x_i^p - 1 \right)$$

which is therefore Schur-convex. Note that the order- p Tsallis entropy is usually defined as $\frac{1}{p-1}(1 - \sum_i x_i^p)$ (Tsallis, 1988) so that our terminology only coincides with the standard one for $p \geq 0$, and differs by a negative sign for $p < 0$. Nonetheless, we are here adopting the convention that the term ‘entropy’ (respectively, ‘negentropy’) should be reserved for functions that are Schur-concave (respectively, Schur-convex).

Applying Theorem 27 (which remains valid under composition with f), we obtain the order- p relative Tsallis

entropy to q (Tsallis, 1998):

$$H_p^{\text{Ts}}(x||q) := \frac{\text{Sgn}(p)}{p - 1} \left(\sum_{i=1}^{d_x} x_i^p q_i^{1-p} - 1 \right).$$

(Again, we deviate from the standard definition by a negative sign for $p < 0$.) Finally, taking q to be the uniform distribution, we obtain what we call the order- p Tsallis nonuniformity:

$$\begin{aligned} I_p^{\text{Ts}}(x) &:= H_p^{\text{Ts}}(x||m^{(d_x)}) \\ &= d_x^{p-1} \left(H_p^{\text{Ts}}(m^{(d_x)}) - H_p^{\text{Ts}}(x) \right). \end{aligned}$$

An interesting fact about $g_p(a)$ is that in the limit $p \rightarrow 1$, it converges to $a \ln a$, so that the order- p Tsallis nonuniformity converges to the Shannon nonuniformity (up to the multiplicative factor $\ln 2$) in the limit $p \rightarrow 1$. The $p \rightarrow 0^+$ and $p \rightarrow \infty$ limits are also important because they are parent quantities of nonuniformity monotones based on the max- and min-entropies, which will be discussed shortly.

Repeating the construction above, with g_p as in (4.9), but alternative choice of function $f(t) := \text{Sgn}(p) \log(\pm t)/(p - 1)$, we obtain the order- p Rényi negentropy

$$-H_p(x) := \frac{\text{Sgn}(p)}{p - 1} \log \sum_{i=1}^{d_x} x_i^p,$$

which is again Schur-convex. Moving on, we construct the order- p relative Rényi entropy of x to q ,

$$H_p(x||q) := \frac{\text{Sgn}(p)}{p - 1} \log \sum_{i=1}^{d_x} x_i^p q_i^{1-p},$$

and the nonuniformity monotone

$$I_p(x) := H_p(x||m^{(d_x)}) = \text{Sgn}(p) \log d_x - H_p(x)$$

which we call the order- p Rényi nonuniformity of x .

Although the functions $-H_p$, $H_p(\cdot||q)$ and I_p are not defined at $p \in \{0, 1\}$, they do converge in these limits and in the limits $p \rightarrow \pm\infty$.

The $p \rightarrow 1$ limit is the least interesting: it just yields the functions H , $H(\cdot||q)$ and I (the Shannon negentropy, Shannon relative entropy and Shannon nonuniformity).

The limits $p \rightarrow 0^+$ and $p \rightarrow \infty$ are of special significance. They will reappear below in different contexts, and are related to the min- and max-entropies. Since there are different versions of these entropies in the literature, we spell out the details in formal definitions. At this point, we are using the definitions by Renner (Renner, 2008), but we will later depart from his conventions in the case of smoothed entropies.

The limit $p \rightarrow 0^+$ yields the Schur-convex function

$$-H_0(x) := -\log |\text{supp}(x)|,$$

which is the negative of what is traditionally called the *max-entropy* of x (because $H_0(x)$ attains the maximum value among the $H_p(x)$) (Gorban *et al.*, 2010). Hence we call this function the *neg-max-entropy*. The corresponding Schur-convex function relative to q is called the *relative max-entropy of x to q* ¹³:

$$H_0(x||q) := -\log \left(\sum_{i \in \text{supp}(x)} q_i \right),$$

whence the resulting nonuniformity monotone is

$$\begin{aligned} I_0(x) &:= H_0 \left(x \middle\| m^{(d_x)} \right) \\ &= \log d_x - \log |\text{supp}(x)| \\ &= \log d_x - H_0(x). \end{aligned} \quad (4.10)$$

The limit $p \rightarrow \infty$, on the other hand, yields the Schur-convex function

$$\begin{aligned} -H_\infty(x) &:= \log \max_{i \in \{1, \dots, d_x\}} x_i \\ &= \log \left(x_1^\downarrow \right), \end{aligned} \quad (4.11)$$

which is the negative of what is traditionally known as the *min-entropy* of x , and which we term the *neg-min-entropy*. The corresponding Schur-convex function relative to q is what is called the *relative max-entropy of x to q* (Datta, 2009):

$$H_\infty(x||q) := \log \max_{i \in \{1, \dots, d_x\}} \left(\frac{x_i}{q_i} \right),$$

whence the resulting nonuniformity monotone is

$$\begin{aligned} I_\infty(x) &:= H_\infty \left(x \middle\| m^{(d_x)} \right) \\ &= \log d_x + \log \left(x_1^\downarrow \right) \\ &= \log d_x - H_\infty(x). \end{aligned} \quad (4.12)$$

Finally, consider the limit $p \rightarrow -\infty$. Here we find that

$$\begin{aligned} -H_{-\infty}(x) &:= -\log \min_{i \in \{1, \dots, d_x\}} x_i, \\ H_{-\infty}(x||q) &:= -\log \min_{i \in \{1, \dots, d_x\}} \left(\frac{x_i}{q_i} \right), \\ I_{-\infty}(x) &:= H_{-\infty} \left(x \middle\| m^{(d_x)} \right) \\ &= -\log d_x - H_{-\infty}(x). \end{aligned}$$

It is straightforward to verify that

$$H_{-\infty}(x||q) = H_\infty(q||x).$$

¹³ It has also been called the min-relative entropy in (Datta, 2009) because among relative entropies, it has the minimum value.

Remark 28 $H_{-\infty}(x||q)$ admits of a simple operational interpretation: $2^{-H_{-\infty}(x||q)}$ is the maximum probability with which the state q can appear in a mixture of states that yields x ¹⁴. The proof is straightforward: if the state q appears with weight w in a mixture of states that yields x , then $x = wq + (1-w)z$ for some state z . In this case, all the components of $x - wq$ must be positive. Consequently,

$$w \leq \min_{i \in \{1, \dots, d_x\}} \left(\frac{x_i}{q_i} \right) = 2^{-H_{-\infty}(x||q)}.$$

Because the order- $(-\infty)$ Rényi nonuniformity, $I_{-\infty}$, is obtained by setting $q = m$ in $H_{-\infty}(x||q)$, it follows that this monotone quantifies the largest probability with which the uniform state m can appear in a convex decomposition of x .¹⁵

The Rényi p -nonuniformities will have an important role to play in the discussion of state conversion with a catalyst. In anticipation of those results, we note two facts about these monotones. First, they are all additive, that is,

$$I_p(x \otimes y) = I_p(x) + I_p(y) \quad (4.13)$$

for all $p \in \mathbb{R}$. This follows from the simple identity

$$\sum_{i,j} (x_i y_j)^p = \left(\sum_i x_i^p \right) \left(\sum_j y_j^p \right).$$

Secondly, the different Rényi entropies are ordered as follows:

$$I_{-\infty}(x) \geq I_{p'}(x) \geq I_p(x), \quad p' < p < 0, \quad (4.14)$$

$$I_0(x) \leq I_p(x) \leq I_{p'}(x) \leq I_\infty(x), \quad 0 < p < p'. \quad (4.15)$$

Inequalities (4.15) follow from a similar order relation over the p -norms $\|x\|_p \equiv (\sum_i x_i^p)^{1/p}$, while (4.14) follow from Lemma 17 of (Brandão *et al.*, 2015). There, it is proven that for any distribution x with full support, $H_p(x)$, considered as a function of p , is nondecreasing over the range $p < 0$, whereas for any x whose support is not the full sample space, $H_p(x) \rightarrow -\infty$ for all $p < 0$. Consider first the case of an x with full support. We find that

$$\begin{aligned} \lim_{p \rightarrow 0^-} H_p(x) &= -\log d_x; \\ \lim_{p \rightarrow 0^+} H_p(x) &=: H_0(x) = \log d_x. \end{aligned}$$

¹⁴ The analogue of this property for the quantum version of this quantity has been noted in (Datta, 2013).

¹⁵ As such, it is analogous to the one from entanglement theory called “the separability of a state” and which is defined as the maximum probability with which a separable state appears in a convex decomposition of the state (Lewenstein and Sanpera, 1998).

Therefore,

$$\lim_{p \rightarrow 0^-} I_p(x) = -\log d_x - \lim_{p \rightarrow 0^-} H_p(x) = 0;$$

$$I_0(x) = \log d_x - H_0(x) = 0.$$

On the other hand, for some x whose support is a proper subset of the sample space,

$$\lim_{p \rightarrow 0^-} I_p(x) = -\log d_x - \lim_{p \rightarrow 0^-} H_p(x) = \infty;$$

$$I_0(x) = \log d_x - H_0(x) = \log \left(\frac{d_x}{|\text{supp}(x)|} \right).$$

It follows from the inequalities in (4.14), (4.15) that for any x (fully-supported or otherwise),

$$I_0(x) = \inf_{p \in \mathbb{R}} I_p(x)$$

$$I_\infty(x) = \sup_{p \in \mathbb{R}_+} I_p(x). \quad (4.16)$$

See also (van Erven and Harremoës, 2010) for a treatment of the connection of the Rényi relative entropies with majorization and Lorenz curves.

B. Nonuniformity monotones arising from the geometry of the Lorenz curve

It is also interesting to consider nonuniformity monotones that are inspired by features of the Lorenz curve. Lorenz curves were originally introduced in economics to characterize income inequality (MOA). There are a number of measures of income inequality that were defined in terms of the Lorenz curve. These immediately yield interesting nonuniformity monotones.

The *Gini index* of x is the area of the Lorenz curve of x over the diagonal (Gini, 1912)¹⁶ (see also F.4.a on p. 563 of MOA). This is clearly a monotone because if the Lorenz curve of x is nowhere below the Lorenz curve of y , then the area over the diagonal of the Lorenz curve of x is not less than that of y . A straightforward calculation yields the value of the Gini index of x in terms of its components:

$$M_{\text{Gini}}(x) = \frac{d_x - 1}{2d_x} - \frac{1}{d_x} \sum_{i=2}^{d_x} (i-1)x_i^\downarrow.$$

The *Schutz index* of x is the maximum vertical deviation between the Lorenz curve of x and the line joining (0,0) to (1,1) (Schutz, 1951) (see also F.4.f on p. 565 of MOA). If the Lorenz curve of x is everywhere above the Lorenz curve of y , the vertical deviation for y at any abscissa is no greater than that of x at the same abscissa. Consequently, the same holds for the maximum vertical

deviation. By virtue of the monotonicity of the vertical deviation over any of the linear segments of the Lorenz curve of x , the Schutz index of x is given in terms of its components as

$$M_{\text{Schutz}}(x) = \max_{k \in \{1, \dots, d_x\}} \left(\sum_{i=1}^k x_i^\downarrow - \frac{k}{d_x} \right).$$

The length of the Lorenz curve of x is also a nonuniformity monotone. In the context of income inequality, it was proposed in (Amato, 1968) and later in (Kakwani, 1980) (see also F.4.h on p. 565 of MOA). We will call it the *Amato index*. It is seen to be a monotone by virtue of the fact that the Lorenz curves of x and y have the same boundary points and are convex, so that if the Lorenz curve of x is nowhere below the Lorenz curve of y , it cannot be shorter. The Amato index of x is given by

$$M_{\text{Amato}}(x) = \sum_{i=1}^{d_x} \sqrt{\frac{1}{d_x^2} + x_i^2}.$$

It was noted in Section IV.A that this monotone can be constructed from the convex function $g(a) = \sqrt{1 + a^2}$.

The nonuniformity monotones corresponding to various limits of the family of Rényi nonuniformities, specifically I_∞, I_0 and $I_{-\infty}$, have very simple interpretations in terms of the geometry of the Lorenz curve. Indeed, this connection allows us to deduce that they are nonuniformity monotones merely from the characterization of noisy-majorization in terms of one Lorenz curve being everywhere not less than another.

For any Lorenz curve L_x , define the *on-ramp* to be the first segment of the curve. Its slope is clearly determined by the largest eigenvalue of x , x_1^\downarrow . Specifically, the slope of the on-ramp is $m^{\text{on}}(x) := x_1^\downarrow / (1/d_x)$. Recalling the definition of the order- (∞) Rényi nonuniformity, Eq. (4.12), we conclude that

$$m^{\text{on}}(x) = 2^{I_\infty(x)}. \quad (4.17)$$

If x is mapped to y by noisy operations, the Lorenz curve L_y must lie on or below L_x , and consequently, the slope of the on-ramp of L_y must be less than or equal to that of L_x . This is an intuitive way to see why I_∞ is a monotone.

Similarly, the *tail* of a Lorenz curve L_x , i.e. the rightmost part where the curve is flat and attains the value one, is related to the number of zero eigenvalues in x . The length $\ell(x)$ of this tail is $\ell(x) = \frac{1}{d_x} (d_x - |\text{supp}(x)|)$, which by Eq. (4.10), implies that

$$\ell(x) = 1 - 2^{-I_0(x)}. \quad (4.18)$$

Again, if x is mapped to y by noisy operations, the Lorenz curve L_y must lie on or below L_x , and this is possible only if the tail lengths satisfy $\ell(y) \leq \ell(x)$. This is a simple way of seeing why I_0 is a nonuniformity monotone.

¹⁶ In fact, the Gini index is usually defined as twice this area.

A sampling of nonuniformity monotones derived from the geometry of the Lorenz curve

Geometric feature of Lorenz curve	Nonuniformity monotone	Comments
Height at point u $u \in [0, 1]$	$M_{\text{Lorenz},u}(x) := L_x(u) = S_{\lfloor d_x u \rfloor}(x)$ $+ [S_{\lfloor d_x u \rfloor + 1}(x) - S_{\lfloor d_x u \rfloor}(x)](d_x u - \lfloor d_x u \rfloor)$ where $S_k(x) := \sum_{i=1}^k x_i^\downarrow$	
Area above diagonal	$M_{\text{Gini}}(x) = \frac{d_x - 1}{2d_x} - \frac{1}{d_x} \sum_{i=2}^{d_x} (i-1)x_i^\downarrow$	Gini index
Maximum vertical deviation from diagonal	$M_{\text{Schutz}}(x) = \max_{k \in \{1, \dots, d_x\}} \left(\sum_{i=1}^k x_i^\downarrow - \frac{k}{d_x} \right)$	Schutz index
Length of Lorenz curve	$M_{\text{Amato}}(x) = \sum_{i=1}^{d_x} \sqrt{\frac{1}{d_x^2} + x_i^2}$	Amato index
Slope of on-ramp	$m^{\text{on}}(x) = 2^{I_\infty(x)}$	$I_\infty(x)$ is the order- (∞) Renyi nonuniformity
Length of tail	$\ell(x) = 1 - 2^{-I_0(x)}$	$I_0(x)$ is the order- (0) Renyi nonuniformity
Negative of slope of off-ramp	$-m^{\text{off}}(x) = -2^{-I_{-\infty}(x)}$	$I_{-\infty}(x)$ is the order- $(-\infty)$ Renyi nonuniformity

TABLE II The first column specifies a geometric property of the Lorenz curve of the classical distribution x . The second specifies the classical nonuniformity monotone that this property defines.

Finally, we mention another monotone that is given by the geometry of the Lorenz curve. Defining the *off-ramp* of the curve to be the final segment — the one that touches the $(1,1)$ point — the monotone is the slope of the off-ramp. This is zero if the Lorenz curve has a nontrivial tail and is nonzero otherwise.

This slope is given by $m^{\text{off}}(x) = 2^{\log d_x + \log x_{d_x}^\downarrow}$ where $x_{d_x}^\downarrow$ is its *smallest* component of x (recall that the min-entropy is defined in terms of the *largest* component of x). Recalling the definition of $I_{-\infty}(x)$, the order- $(-\infty)$ Renyi nonuniformity, Eq. (4.12), we conclude that

$$m^{\text{off}}(x) = 2^{-I_{-\infty}(x)}. \quad (4.19)$$

Clearly, if the Lorenz curve of x is nowhere below that of y , the off-ramp slope of x is less than or equal to that of y . Hence the negative of the off-ramp slope is a nonuniformity monotone, which is a simple way of seeing that $I_{-\infty}(x)$ is a monotone.

C. Nonuniformity monotones from other Schur-convex functions

Not all Schur-convex functions lend themselves to a decomposition as the sum of a single convex function evaluated on the components, but we can still construct nonuniformity monotones from some of these.

Proposition 29 *Given a distribution, $q \in \mathbb{R}_+^{d_q}$, and a non-negative function $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, the function $\Gamma(\cdot \| q) : \mathbb{R}_+^{d_q} \rightarrow \mathbb{R}$ defined by*

$$\Gamma(x \| q) = \prod_{j=1}^{d_q} \left[g \left(\frac{x_j}{q_j} \right) \right]^{q_j} \quad (4.20)$$

is Schur-convex (Schur-concave) relative to q if $\log g$ is convex (concave).

Proof. By taking the log on both sides of (4.20) and using Lemma 25 we get that $\log(\Gamma(x \| q))$ is Schur-convex. Since the log function is monotonically increasing we get that $\Gamma(x \| q)$ must also be Schur-convex. The concave case follows the same lines. ■

As an example, consider the function $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ defined by $g(a) = a$. Since $\log a$ is concave, the proposition above implies that the function $\prod_{j=1}^{d_q} \left(\frac{x_j}{q_j} \right)^{q_j}$ is Schur-concave relative to q and therefore

$$\Gamma_{WG}(x \| q) = 1 - \prod_{j=1}^{d_q} \left(\frac{x_j}{q_j} \right)^{q_j}$$

is Schur-convex relative to q . Note that the product is the *weighted* geometric mean (hence ‘WG’) of the elements x_j/q_j with weights q_j .

For $q = m^{(d_x)}$ this example yields the nonuniformity monotone

$$M_G(x) := 1 - d_x \left(\prod_{i=1}^{d_x} x_i \right)^{\frac{1}{d_x}}.$$

Because the weighted geometric mean is uniformly weighted in this case, it is just the geometric mean (hence the subscript ‘G’). It is the analogue for nonuniformity of the G -concurrence entanglement monotone (Gour, 2005).

Note that the product of components, $\prod_{i=1}^{d_x} x_i$, and the geometric mean of the components, $\left(\prod_{i=1}^{d_x} x_i \right)^{\frac{1}{d_x}}$ are both

Schur-concave functions, but they generalize to the same nonuniformity monotone, namely $M_G(x)$, because of the requirement of eq. (4.3).

In fact, the product of all the components is just one among a whole class of Schur-concave functions: the elementary symmetric polynomials in the components. The k th such polynomial is defined as $S_k^{(d)} : \mathbb{R}_+^d \rightarrow \mathbb{R}$,

$$S_k^{(d)}(x) := \sum_{i_1 < i_2 < \dots < i_k} \prod_{j=1}^k x_{i_j}, \quad (4.21)$$

where the sum is over all k -tuples of distinct components. We have not found a way to construct nonuniformity monotones from the Schur-concave function $S_k^{(d)}$ except for the special cases of $k = d$ and $k = 2$ (the $k = 1$ case is trivial).

Finally, we summarize in Table III the operational interpretations of some of these measures, interpretations that will be discussed in the rest of this article.

V. EXACT STATE CONVERSION

A. A standard form of nonuniformity: sharp states

It is useful to define a nonuniform state that can serve as a *natural unit* for the resource in question. For the resource of nonuniformity, the pure state of a two-level system, that is, the distribution $(1, 0)$, is an obvious choice. We refer to such a state as a *bit of nonuniformity* or as a *pure bit*.

The family of pure states of different dimensions also provides a natural standard relative to which one can quantify other states' nonuniformity. A pure state of dimension d corresponds to the distribution

$$(1, \underbrace{0, \dots, 0}_{d-1}).$$

Clearly, k pure bits constitute a pure state of dimension 2^k , however, pure states for dimensions that are not powers of 2 cannot be represented as an integer number of pure bits.

Pure states of different dimensions are themselves subsumed as a special case of a third family of nonuniform states, which we will call the *sharp states*. It is the latter set which is the most versatile and which we will take in this article to be the standard relative to which we judge the nonuniformity of states. A sharp state is any distribution of the form

$$\left(\underbrace{\frac{1}{d_u}, \dots, \frac{1}{d_u}}_{d_u}, 0, \dots, 0 \right), \quad (5.1)$$

$\underbrace{\hspace{10em}}_d$

where $d, d_u \in \mathbb{Z}_+$ and $d_u \leq d$, that is, a state of dimension d that is uniformly distributed over d_u physical states and assigns probability zero to the rest.

If we evaluate the Shannon nonuniformity of such a state, we find $\log(d/d_u)$. Furthermore, one easily verifies that the order-0 and order- ∞ Rényi nonuniformities are equal to the Shannon nonuniformity for this state. It then follows from Eq. (4.16) that for all $p \in \mathbb{R}_+$ the order- p Rényi nonuniformity is equal to the Shannon nonuniformity for this state. Indeed, the noisy equivalence class of this state is the set of all sharp states with the same ratio of d to d_u and therefore on the set of sharp states *every* nonuniformity monotone can be expressed as a function of the Shannon nonuniformity. We will therefore adopt the convention of referring to $\log(d/d_u)$ as simply *the nonuniformity* of the sharp state. Clearly, the possible values for the nonuniformity of a sharp state is just the image of the rationals greater than 1 under the logarithmic function.

We will adopt as our canonical representative of each equivalence class of sharp states the one of lowest dimension within that class, and we will label it by its nonuniformity. The canonical sharp state with nonuniformity I will be denoted s_I . Thus, we have

$$s_0 := (1) \quad (5.2)$$

$$s_{\log(3/2)} := (1/2, 1/2, 0) \quad (5.3)$$

$$s_1 := (1, 0) \quad (5.4)$$

$$s_{\log(7/3)} := (1/3, 1/3, 1/3, 0, 0, 0, 0) \quad (5.5)$$

$$s_{\log 3} := (1, 0, 0) \quad (5.6)$$

$$s_2 := (1, 0, 0, 0) \quad (5.7)$$

Note that a uniform state is a sharp state with nonuniformity zero. The canonical sharp state in this class, s_0 , is the uniform state of smallest dimension, namely, dimension 1 (which we could also denote by $m^{(1)}$) but it is of course noisy-equivalent with the uniform state $m^{(d)}$ of any dimension. A pure bit is the canonical sharp state with nonuniformity 1, s_1 . A pure state of dimension d is the canonical sharp state with nonuniformity $\log d$, $s_{\log d}$.

The Lorenz curves representing s_I for various values of I are depicted in Fig. 4. It is easy to see that the set of sharp states includes all and only those states for which the Lorenz curve has the simple form of an on-ramp and a tail. Recalling Eqs. (4.17) and (4.18), the on-ramp slope of a sharp state is $m^{\text{on}}(s_I) = 2^I$, while the tail length is $\ell(s_I) = 1 - 2^{-I}$. It is clear from the geometry of the Lorenz curves that the set of sharp states are totally ordered under noisy operations. For any degree of nonuniformity that is equal to the logarithm of a real (possibly irrational) number, one can approximate this arbitrarily closely with a sequence of sharp states whose nonuniformity converges to that degree of nonuniformity in the limit of arbitrarily large dimension.

Finally, note that a tensor product of two sharp states, one with nonuniformity I_1 and one with nonuniformity I_2 is another sharp state with nonuniformity $I = I_1 + I_2$, that is, $s_{I_1} \otimes s_{I_2}$ is noisy-equivalent with $s_{I_1+I_2}$.

Nonuniformity monotone	Operational Interpretation	Comments
$I_{-\infty}(x) := -\log d_x - H_{-\infty}(x)$	Quantifies maximum weight of uniform state in decomposition	See Remark 28
$I_0(x) := \log d_x - H_0(x)$	Quantifies single-shot distillable nonuniformity	See Lemma 30
$I_{\infty}(x) := \log d_x - H_{\infty}(x)$	Quantifies single-shot nonuniformity of formation	See Lemma 31
$I(x) := \log d_x - H(x)$	Quantifies asymptotic rate of interconversion	See Lemma 66
	Describes total order of states under approximate catalysis	See Theorem 69

TABLE III Operational interpretation of some nonuniformity monotones

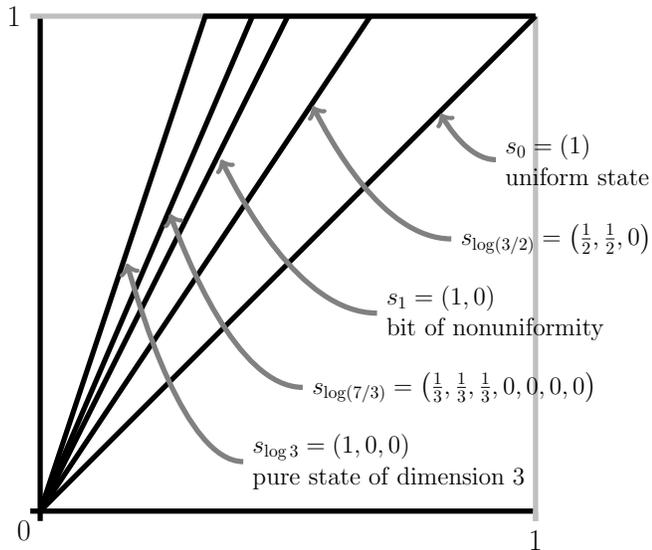


FIG. 4 The Lorenz curves of various sharp states.

B. Nonuniformity of formation and distillable nonuniformity

Given this standard form of nonuniformity, two important questions arise:

- (i) What is the minimum nonuniformity of sharp state required to deterministically create (or “form”) a single copy of state x by noisy operations?
- (ii) Given a single copy of state x , what is the maximum nonuniformity of sharp state that can be deterministically extracted (or “distilled”) from it by noisy operations?

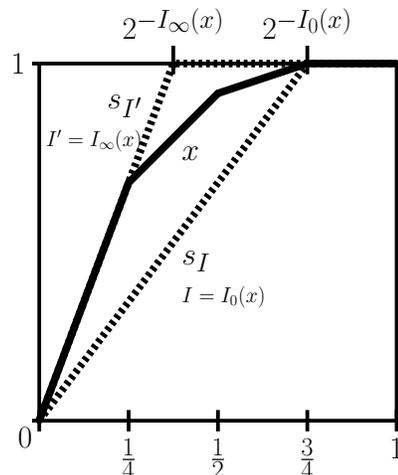
We refer to the answer to the first question as the *single-shot nonuniformity of formation* and the answer to the second question as the *single-shot distillable nonuniformity*. The term “single-shot” refers to our interest in forming or distilling only one copy of x .

If we find that the minimum nonuniformity of any sharp state required to form x is I , then the minimum number of pure bits required is $\lceil I \rceil$, where $\lceil a \rceil$ denotes the smallest integer not smaller than a (the ceiling function). Similarly, if the maximum nonuniformity of any sharp state that can be extracted is I , then the maximum number of pure bits that can be extracted is $\lfloor I \rfloor$,

where $\lfloor a \rfloor$ denotes the largest integer not larger than a (i.e. the floor function).

Note that the possession of a pure bit is equivalent, as a resource, to having access to one implementation of a one-bit erasure operation¹⁷. This follows from the fact that (a) given one pure bit as a resource, one can implement erasure on a system by swapping the system’s state with the resource’s state, and (b) given one implementation of a one-bit erasure operation, a uniform state (available for free) can be transformed into one bit of nonuniformity. It is useful to think of this equivalence class of resources as the ability to do one bit of “informational work”. So distillation of nonuniformity is the analogue, within the resource theory of nonuniformity, of work extraction in the resource theory of athermality. Indeed, work extraction often involves a two-step procedure wherein one first distills pure bits, and then uses them in a Szilard engine to do work (Dahlsten *et al.*, 2009; Egloff *et al.*, 2012).

Questions (i) and (ii) above are easily answered geometrically by means of Lorenz curves, as illustrated in Fig. 5. This proof technique was first used in (Horodecki and Oppenheim, 2013).



¹⁷ It is similarly evident that a sharp state with nonuniformity $\log d$ is equivalent, as a resource, to having access to an operation that takes an arbitrary state of dimension d to a sharp state with nonuniformity $\log d$.

FIG. 5 From the Lorenz curve of x , one easily infers that $I_0(x)$ is the maximum nonuniformity of sharp state that can be distilled from x and that $I_\infty(x)$ is the minimum nonuniformity of sharp state that is required to form x .

For a given state x , its Lorenz curve L_x lies above or on the Lorenz curve L_{s_I} of a sharp state with nonuniformity I if and only if it has at least as long a tail length, i.e., if and only if $\ell(x) \geq \ell(s_I)$. Given the expression for the tail length of x , Eq. (4.18), and the fact that the tail length of s_I is simply $\ell(s_I) = 1 - 2^{-I}$, the condition becomes $I \leq I_0(x)$. Since $I_0(x)$ is the logarithm of a rational, we can attain the bound using a uniform ancilla of finite dimension.

Similarly, there is a simple geometric condition that is necessary and sufficient to guarantee that the Lorenz curve L_{s_I} of a sharp state with nonuniformity I lies on or above the Lorenz curve L_x of x : namely, the on-ramp slope of L_{s_I} must equal or exceed that of L_x , that is, $m_{s_I}^{\text{on}} \geq m_x^{\text{on}}$. According to Eq. (4.17) and the fact that the on-ramp slope of L_{s_I} is 2^I , this is equivalent to $I \geq I_\infty(x)$. Here, we cannot always attain the bound with a finite-dimensional ancilla, since $I_\infty(x)$ can be the logarithm of an irrational number. However, we can approach equality arbitrarily closely by using large enough ancillary uniform states.

We can summarize what we have proven as follows.

Proposition 30 *The state conversion $x \xrightarrow{\text{noisy}} s_I$ is possible if and only if*

$$I \leq I_0(x).$$

Consequently, the maximum nonuniformity of sharp state that can be distilled deterministically from a state x by noisy operations is $I_0(x)$, the order-0 Rényi nonuniformity of x .

Proposition 31 *The state conversion $s_I \xrightarrow{\text{noisy}} x$ is possible if and only if*

$$I \geq I_\infty(x).$$

Consequently, the minimum nonuniformity of sharp state that is required to deterministically form a state x by noisy operations is $I_\infty(x)$, the order-(∞) Rényi nonuniformity of x .

These results yield an operational interpretation of $I_\infty(x)$ ($I_0(x)$) in terms of the single-shot nonuniformity cost (yield) of a state x ¹⁸.

¹⁸ Note that it is not, strictly speaking, the single-shot cost (yield) of pure bits because the latter is the integer ceiling (floor) of this quantity. Nonetheless, as long as one remembers this subtlety, pure bit cost (yield) is an accurate description and it is often described this way (Dahlsten *et al.*, 2009; Del Rio *et al.*, 2011; Faist *et al.*, 2012).

Note that if the system is composite and the state is a product state, then by virtue of the additivity of I_∞ and I_0 , the distillable nonuniformity and the nonuniformity of formation of the composite state are just the sum of those of the components. In other words, the maximum nonuniformity yield and the minimum nonuniformity cost are both achievable by processing the components separately. It also follows that if one cannot distill any nonuniformity from x then one cannot do so from any number of copies of x either. The analogous fact was pointed out for athermality theory in (Janzing *et al.*, 2000).

Also note that if the system consists of several components, correlations among the systems is a resource of nonuniformity. For instance, a pair of bits which are perfectly correlated but with uniform marginals on each bit define a probability distribution $x := (1/2, 0, 0, 1/2)$, corresponding to a sharp state with nonuniformity 1. This is noisy-equivalent to a single pure bit. The point is that, in the resource theory of nonuniformity, all unitaries are free, including those that couple together distinct systems. It follows that how nonuniformity is encoded within a composite system (locally or in correlations) is irrelevant.

In the resource theory of athermality, the problem that is the analogue of finding the maximum single-shot distillable nonuniformity is finding the maximum amount of work that can be extracted from a state in a single-shot protocol. This was first solved in (Horodecki and Oppenheim, 2013) and (Åberg, 2013). That the single-shot distillable nonuniformity is $I_0(x)$ is simply the specialization to energy-degenerate systems of their result. (Horodecki and Oppenheim, 2013) also determined the work cost of preparing a state in a single-shot protocol, which in the case of energy-degenerate systems is just the single-shot nonuniformity of formation, and their result reduces to $I_\infty(x)$ in this case.

Finally, the results on single-shot formation and distillation of nonuniformity yield a simple sufficient condition for the possibility of a state conversion.

Lemma 32 *If x and y are states such that*

$$I_0(x) \geq I_\infty(y),$$

then $x \xrightarrow{\text{noisy}} y$.

Equivalently, the function

$$W(x, y) := I_0(x) - I_\infty(y)$$

is a go witness for the state conversion $x \xrightarrow{\text{noisy}} y$, which is to say that if $W(x, y) \geq 0$ then $x \xrightarrow{\text{noisy}} y$.

Proof. The result follows from Propositions 30 and 31. The latter implies that a sharp state with nonuniformity $I_\infty(y)$ suffices to form y , and the former implies that starting from x , we can distill a sharp state with nonuniformity $I_0(x)$ (recall that $2^{I_0(x)}$ is always rational). Consequently, if $I_0(x) \geq I_\infty(y)$, we can distill from x a sharp state that is sufficient to form y . ■

C. Nonuniformity cost and yield of state conversion

We have discussed the problem of whether or not x can be mapped to y by noisy operations. We have also considered how much nonuniformity (in some standard form) may be extracted from a state, and how much nonuniformity is required to create the state. There is an obvious way in which to combine these questions into another pair of questions:

- (i) If $x \xrightarrow{\text{noisy}} y$, what is the maximum nonuniformity of sharp state that can be distilled in addition to achieving the state conversion?
- (ii) If it is not the case that $x \xrightarrow{\text{noisy}} y$, what is the minimum nonuniformity of sharp state that one requires to make the state conversion possible?

We call the former the *nonuniformity yield* of the state conversion and the latter the *nonuniformity cost* of the state conversion.

The following lemma will be useful for answering these questions.

Lemma 33 *The operation of adjoining an ancillary system in sharp state s_I to a system in state x , that is, $x \mapsto x \otimes s_I$, corresponds to the map $L_x(u) \mapsto L_{x \otimes s_I}(u)$ on Lorenz curves, where*

$$L_{x \otimes s_I}(u) = \begin{cases} L_x(2^I u), & u \in [0, 2^{-I}] \\ 1, & u \in (2^{-I}, 1] \end{cases} \quad (5.8)$$

This corresponds to a 2^I -fold compression of the Lorenz curve of x along the u -axis while lengthening the tail an appropriate amount.

To see this, note that if $I = \log(d/d_u)$, then

$$(y \otimes s_I)^\downarrow = \underbrace{(y_1^\downarrow/d_u, \dots, y_1^\downarrow/d_u)}_{d_u}, \dots, \underbrace{(y_{d_u}^\downarrow/d_u, \dots, y_{d_u}^\downarrow/d_u)}_{d_u}, \underbrace{(0, \dots, 0)}_{d_y(d-d_u)}, \quad (5.9)$$

where the multiplicity of each nonzero coefficient is d_u and the multiplicity of the zeros is $d_y(d-d_u)$. One then verifies that the Lorenz curve has the form described.

The case of $I = 0$ corresponds to adjoining an ancillary system in the uniform state, which leaves the Lorenz curve invariant.

It is also useful to define the following order relation on states.

Definition 34 (λ -noisy-majorization) *We say that x λ -noisy-majorizes y , where $\lambda \in \mathbb{R}$, if*

$$L_x(u) \geq L_y(2^\lambda u) \quad \forall u \in [0, u_*], \quad (5.10)$$

where $u_* := \min\{1, 2^{-\lambda}\}$, or equivalently,

$$L_x(2^{-\lambda} u) \geq L_y(u) \quad \forall u \in [0, u_*^{-1}]. \quad (5.11)$$

We will seek to understand this relation in terms of the geometry of Lorenz curves. We consider the cases of $\lambda > 0$ and $\lambda < 0$ in turn.

For $\lambda > 0$, if x λ -noisy-majorizes y then *not only* is the Lorenz curve of x everywhere greater than or equal to the Lorenz curve of y , but it is also the case that one can implement a 2^λ -fold compression of the Lorenz curve of y along the u -axis (making it rise to 1 more rapidly), and *still* find that the Lorenz curve of x is everywhere greater than or equal to it. Equivalently, it asserts that one can implement a 2^λ -fold *stretching* of the Lorenz curve of x along the u -axis (making it rise to 1 less rapidly), and still find that the Lorenz curve of x is everywhere greater than or equal to that of y .

From Lemma 33 one can also deduce the physical significance of the fact that x λ -noisy-majorizes y when $\lambda > 0$: it implies that $x \xrightarrow{\text{noisy}} y \otimes s_\lambda$, that is, one can distill a sharp state of nonuniformity λ in addition to achieving the conversion of x to y . To see this, note that Eq. (5.10) implies, via Eq. (5.8), that $L_x(u) \geq L_{y \otimes s_\lambda}(u)$ in the region $u \in [0, u_*]$ where $u_* = 2^{-\lambda}$. The only subtlety then, is to prove that $L_x(u) \geq L_{y \otimes s_\lambda}(u)$ also in the region $u \in (u_*, 1]$. This follows from noting that at $u = u_*$, we have $L_{y \otimes s_\lambda}(u_*) = 1$, but we also have $L_x(u_*) \geq L_{y \otimes s_I}(u_*)$, and therefore $L_x(u_*) = 1$. But because the Lorenz curve of x is concave it follows that $L_x(u) = 1$ for $u \in (u_*, 1]$ and consequently it cannot be smaller than $L_y(u)$ in that region.

We now consider the case of $\lambda < 0$. In this case, if x λ -noisy-majorizes y then the Lorenz curve of x must somewhere be lower than that of y (so that x does *not* noisy-majorize y), but if we implement a 2^λ -fold compression of the Lorenz curve of y along the u -axis, which, because $2^\lambda < 1$ is a net *stretching* of the Lorenz curve of y along the u -axis (making it rise to 1 less rapidly), then the Lorenz curve of x becomes everywhere greater than or equal to that of y . Equivalently, it asserts that a $2^{|\lambda|}$ -fold net compression of the Lorenz curve of x along the u -axis (making it rise to 1 more rapidly) can make it everywhere greater than or equal to that of y .

Physically, if x λ -noisy-majorizes y for $\lambda < 0$, then $x \otimes s_{|\lambda|} \xrightarrow{\text{noisy}} y$, that is, a sharp state of nonuniformity $|\lambda|$ can make the conversion of x to y possible.

This time, the proof begins with Eq. (5.11), which implies, via Eq. (5.8), that $L_{x \otimes s_{|\lambda|}}(u) \geq L_y(u)$ in the region $u \in [0, u_*^{-1}]$ where $u_*^{-1} = 2^\lambda$. Again, one can easily infer that this inequality also holds in the region $u \in (u_*^{-1}, 1]$, and so holds for all $u \in [0, 1]$.

Standard noisy-majorization is λ -noisy-majorization with $\lambda = 0$.

Note that the sign of λ determines whether λ -noisy-majorization is a strengthening or a weakening of noisy-majorization: if x λ -noisy-majorize y for a strictly positive λ , then it has *more* nonuniformity than is required for it to noisy-majorize y , while if x λ -noisy-majorizes y for negative λ , then it has *less* nonuniformity than is required for it to noisy-majorize y .

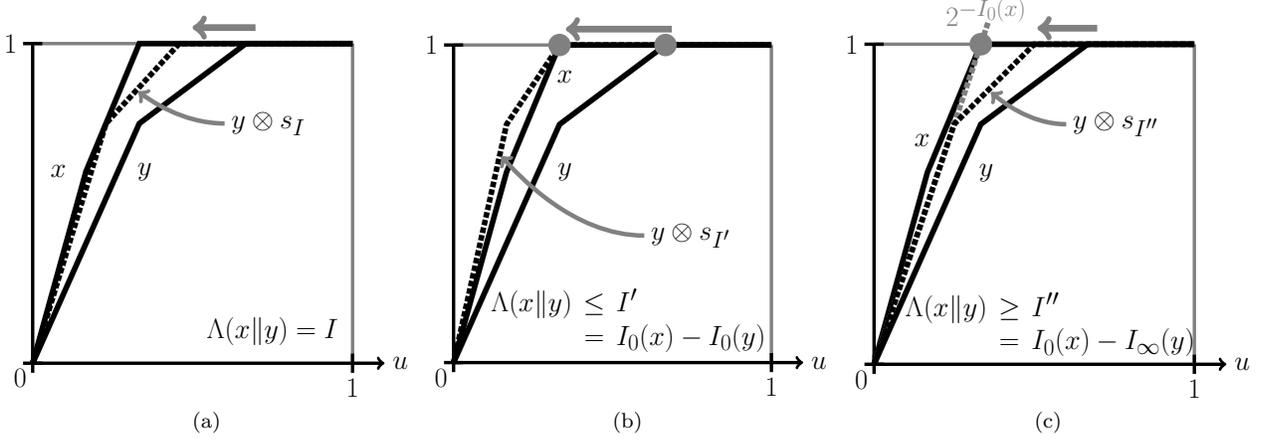


FIG. 6 An illustration of the nonuniformity yield of state conversion ($\Lambda(x||y)$ when $\Lambda(x||y) \geq 0$) and bounds thereon. The Lorenz curves of x and y are depicted, as well as the Lorenz curve of $y \otimes s_I$ (dashed) for various values of I . According to (5.8), the curve of $y \otimes s_I$ is obtained by linearly compressing the curve of y along the u -axis. The nonuniformity yield is the largest value of I such that $x \mapsto y \otimes s_I$, i.e. such that this linearly compressed curve is still below or on the curve of x . This is illustrated in (6a): we compress the curve of y until it touches the curve of x , and the resulting value of I equals the nonuniformity yield $I = \Lambda(x||y)$. In (6b), it is shown how the upper bound $I' := I_0(x) - I_0(y) \geq \Lambda(x||y)$ from Proposition 38 can be derived graphically: compress the Lorenz curve of y such that the leftmost point on its tail is brought to lie on top of the leftmost point on the tail of the Lorenz curve of x . It is clear that the curve L_y cannot be compressed any further without crossing L_x , which gives an upper bound on $\Lambda(x||y)$; in our example, the upper bound is strict, since $L_{y \otimes s_{I'}}$ already crosses L_x . As a curve's tail is related to I_0 as shown in Fig. 5, this graphical observation gives the desired upper bound by simple algebra. By considering the amount of compression of L_y that keeps the on-ramp slope of the resulting Lorenz curve less than that of L_x , we can obtain the upper bound $I_\infty(x) - I_\infty(y)$ (not shown). Fig. (6c) provides a graphical proof of the lower bound $I_0(x) - I_\infty(y) \leq \Lambda(x||y)$. The grey dotted line connects the origin with the leftmost point on the tail of L_x ; it defines the on-ramp of a sharp state of nonuniformity $I_0(x)$ that can be distilled from x . Then $I_\infty(y)$ bits of nonuniformity are spent to form y ; graphically, the curve of y is compressed until its on-ramp slope agrees with that of the sharp state.

Finally, we require a definition before formalizing the answers to the questions that headed this section.

Definition 35 The maximum factor λ by which x λ -noisy-majorizes y is denoted $\Lambda(x||y)$,

$$\Lambda(x||y) := \max \{ \lambda : L_x(u) \geq L_y(2^\lambda u) \quad \forall u \in (0, u_*] \}, \quad (5.12)$$

where $u_* := \min\{1, 2^{-\lambda}\}$.

Given that $x \xrightarrow{\text{noisy}} y \otimes s_I$ if and only if x λ -noisy-majorizes y by a factor $\lambda = I$, it is clear that if the maximum such factor, $\Lambda(x||y)$, is positive, then it quantifies the maximum nonuniformity of sharp state that can be distilled while achieving the state conversion $x \mapsto y$.

Similarly, given that $x \otimes s_I \xrightarrow{\text{noisy}} y$ if and only if x λ -noisy-majorizes y by a factor $\lambda = -I$, it is clear that if the maximum such factor, $\Lambda(x||y)$, is negative, then it quantifies the minimum nonuniformity of sharp state that is required to achieve the state conversion $x \mapsto y$.

To summarize:

Proposition 36 If $\Lambda(x||y) > 0$, then the state conversion $x \xrightarrow{\text{noisy}} y \otimes s_I$ is possible if and only if

$$I \leq \Lambda(x||y).$$

That is, one can distill a sharp state of nonuniformity at most $\Lambda(x||y)$ in addition to achieving the conversion of x to y .

Proposition 37 If $\Lambda(x||y) < 0$, then the state conversion $x \otimes s_I \xrightarrow{\text{noisy}} y$ is possible if and only if

$$I \geq -\Lambda(x||y).$$

That is, in order to achieve the conversion of x to y it costs a sharp state of nonuniformity at least $|\Lambda(x||y)|$.

In general, there is not a simple expression for $\Lambda(x||y)$ because it depends on the details of the shape of the Lorenz curves of x and y .

Nonetheless, one can compute $\Lambda(x||y)$ by finite means. For every height h in the plot of the Lorenz curves of x and y , one obtains a bound on $\Lambda(x||y)$ in terms of the ratio $L_y^{-1}(h)/L_x^{-1}(h)$. However, because $L_y(u)$ and $L_x(u)$ are concave, it suffices to compute these ratios only at the heights h corresponding to the elbows of $L_x(u)$ and $L_y(u)$. Therefore, there are at most $d_x + d_y - 1$ such comparisons that need to be made. It follows that one can define $\Lambda(x||y)$ by

$$\Lambda(x||y) := \max \{ \lambda : L_x(u) \geq L_y(2^\lambda u) \quad \forall u \in \mathcal{U} \}, \quad (5.13)$$

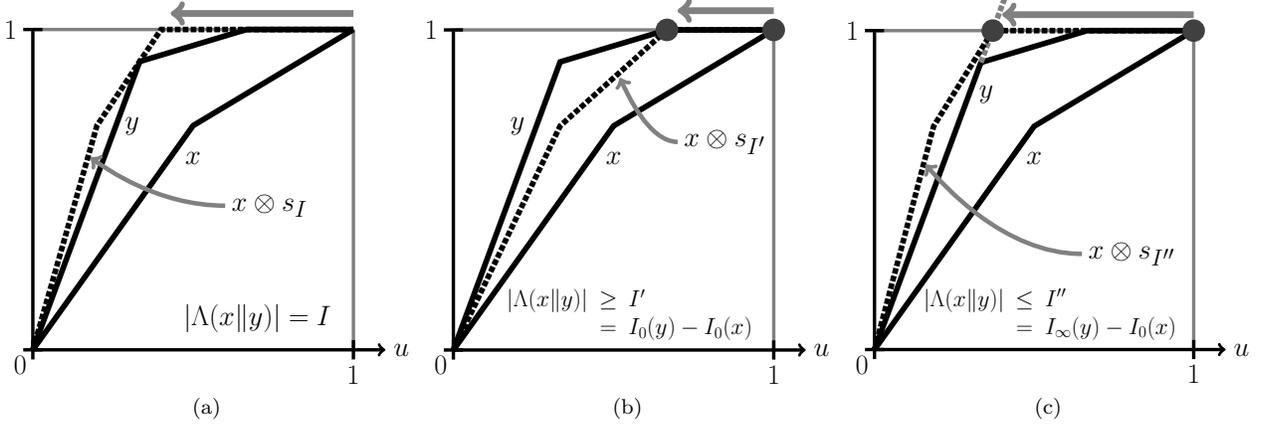


FIG. 7 An illustration of the nonuniformity cost of state conversion ($|\Lambda(x||y)|$ when $\Lambda(x||y) < 0$) and bounds thereon. The Lorenz curves of x and y are depicted, as well as the Lorenz curve of $x \otimes s_I$ (dashed) for various values of I . The nonuniformity cost is the least value of I such that $x \otimes s_I \mapsto y$, which is graphically related to the least amount of compression in the u -direction which is necessary such that the compressed curve of x is on or above the curve of y . This is illustrated in Fig. (7a). The value of I corresponding to this least amount of compression defines the nonuniformity cost of converting x to y , and the quantity $\Lambda(x||y)$ is defined as the negative of this cost, $I = |\Lambda(x||y)|$. Fig. (7b) provides a graphical demonstration of the lower bound on the nonuniformity cost $|\Lambda(x||y)| \geq I_0(y) - I_0(x)$. The curve of x has to be compressed by a factor at least equal to the one that makes the tail of the resulting Lorenz curve as long as the tail of the Lorenz curve of y . In other words, the leftmost point of the tail of L_x has to be brought to lie on top of the leftmost point of the tail of L_y (these points are indicated by grey dots). Considering the compression of L_x required to make the on-ramp slope of the resulting Lorenz curve at least as great as that of L_y yields the lower bound $I_\infty(y) - I_\infty(x)$ (not shown). Fig. (7c) provides a graphical demonstration of the upper bound on the nonuniformity cost $|\Lambda(x||y)| \leq I_\infty(y) - I_0(x)$. The grey dotted line extends the on-ramp of L_y , and its intersection with the horizontal line of height one is indicated by a gray dot. The leftmost point of the tail of L_x is also indicated by a grey dot. If the Lorenz curve of x is compressed to such an extent that the second dot comes to lie on top of the first, then the new Lorenz curve $x \otimes s_I$ is clearly everywhere on or above the curve of y .

where $\mathcal{U} := \left(\left\{ \frac{1}{d_x}, \frac{2}{d_x}, \dots, 1 \right\} \cup \left\{ \frac{1}{d_y}, \frac{2}{d_y}, \dots, 1 \right\} \right) \cap (0, u_*]$ with $u_* := \min\{1, 2^{-\lambda}\}$.

In addition, even without solving the optimization problem, one can determine some nontrivial bounds on $\Lambda(x||y)$.

Proposition 38 $\Lambda(x||y)$ is bounded above and below as follows:

$$I_0(x) - I_\infty(y) \leq \Lambda(x||y), \quad (5.14)$$

$$\Lambda(x||y) \leq \min\{I_0(x) - I_0(y), I_\infty(x) - I_\infty(y)\}. \quad (5.15)$$

If x is a sharp state then

$$\Lambda(x||y) = I(x) - I_\infty(y).$$

If y is a sharp state then

$$\Lambda(x||y) = I_0(x) - I(y).$$

If both x and y are sharp states then

$$\Lambda(x||y) = I(x) - I(y).$$

To see the proof of these bounds, we consider the cases of $\Lambda(x||y) \geq 0$ and $\Lambda(x||y) < 0$ separately, and use intuitions concerning nonuniformity cost and yield of state

conversions. See Figs. 6 and 7 for the associated Lorenz curves.

Consider the case $\Lambda(x||y) \geq 0$ for instance. The lower bound $\Lambda(x||y) \geq I_0(x) - I_\infty(y)$ is evident from the fact that one can always achieve a yield which is at least the excess of the single-shot distillable nonuniformity of x , $I_0(x)$, over the single-shot nonuniformity of formation of y , $I_\infty(y)$. We obtain the *upper bounds* on the nonuniformity yield from the monotonicity and additivity of I_∞ and of I_0 . For instance, we have $I_0(x) \geq I_0(y \otimes s_I) = I_0(y) + I$, which implies that $I \leq I_0(x) - I_0(y)$. Similar arguments hold for the case of $\Lambda(x||y) < 0$.

Finally, recalling that for a sharp state s , $I(s) = I_0(s) = I_\infty(s)$, it follows that if either x or y or both are sharp states, the upper and lower bounds coincide and consequently we obtain an exact expression for Λ .

Note that if y is a uniform state m (which is a sharp state of nonuniformity 0), then $\Lambda(x||y) = \Lambda(x||m) = I_0(x)$, which is positive, so the result predicts a nonuniformity yield of $I_0(x)$. Given that one can prepare m for free, the nonuniformity yield of this state conversion problem is just the nonuniformity that can be distilled from x . This is indeed $I_0(x)$, as demonstrated in Lemma 30.

Similarly, if x is a uniform state m , then $\Lambda(x||y) = \Lambda(m||y) = -I_\infty(y)$, which is negative, so the result predicts a nonuniformity cost of $I_\infty(y)$. Given that m is free,

this state conversion problem is just the problem of forming the state y . This reproduces the result of Lemma 31 that the nonuniformity of formation is $I_\infty(y)$.

A state conversion process that has been of particular interest in the literature is that of erasure, which for a system of dimension d is defined as the process which takes an arbitrary state x to a pure state of dimension d . Because such a pure state is a sharp state with nonuniformity $\log d$, we can immediately infer from our result that the nonuniformity cost of erasure of a state x of a d -dimensional system is

$$\log d - I_0(x) = H_0(x),$$

the Rényi 0-entropy of x . Note that in the athermality theory, where the free states are thermal states at temperature T , the work cost of erasure is simply the nonuniformity cost multiplied by $kT \ln 2$. And hence, the work cost of erasure of a state x is $kT \ln 2 H_0(x)$. In particular, if x is the uniform state of a binary variable, $m^{(2)}$, then the work cost is $kT \ln 2$. This is a form of Landauer's principle (Landauer, 1961). Similarly, pure states can be used to do work in the athermality theory. Bennett was perhaps the first to suggest that an initialized memory tape was capable of storing work (Bennett, 1982).

If we imagine having a reservoir in a sharp state of arbitrarily large nonuniformity, then for any pair of states x and y , one can ask: what is the minimum value of $\lambda_2 - \lambda_1$ such that $x \otimes s_{\lambda_1} \xrightarrow{\text{noisy}} y \otimes s_{\lambda_2}$? If this value is positive, there is a yield of nonuniformity, while if it is negative, there is a cost. As it turns out, if this conversion is possible, then either it is possible with $\lambda_1 = 0$ or it is possible with $\lambda_2 = 0$. The presence of the sharp state reservoir does not increase the yield or reduce the cost of any state conversion. This is a consequence of the uselessness of sharp states as catalysts, Proposition 41 below.

For any given state y , if we consider $\Lambda(x||y)$ as a function over states x , it is a nonuniformity monotone. Similarly, for any given state x , if we consider $-\Lambda(x||y)$ as a function over states y , it is a nonuniformity monotone.

The idea of quantifying “how much” one state majorizes another is from (Egloff *et al.*, 2012). The maximum λ for which x λ -noisy-majorizes y , which we have denoted $\Lambda(x||y)$, was there called the “relative mixedness”. The order relation that we have called λ -noisy-majorization was first introduced in (Faist *et al.*, 2012) as simply “ λ -majorization”. We use the term “noisy-majorization” because we are reserving the term “majorization” for the case where the states are of the same dimension, to accord with standard usage. The identification of the nonuniformity yield of the state conversion $x \mapsto y$ as $\Lambda(x||y)$ was also made in (Faist *et al.*, 2012). This yield translates, in the athermality theory, to work extraction, featured in (Egloff *et al.*, 2012; Skrzypczyk *et al.*, 2013, 2014), and to cooling, featured in (Janzing *et al.*, 2000). That $\Lambda(x||y)$ simplifies to the nonuniformity of formation when x is a uniform state was noted in (Egloff *et al.*, 2012). That paper also presents the

athermality cost of a particular erasure. The upper and lower bounds on the nonuniformity cost of state conversion, together with the exact expressions in the case where one of the states is sharp, have not previously been noted.

Finally, it is worth noting that these results provide necessary and sufficient conditions for a state conversion to be possible.

Proposition 39 *The maximum factor λ such that x λ -noisy-majorizes y , denoted $\Lambda(x||y)$, is a complete witness for the state conversion $x \xrightarrow{\text{noisy}} y$, that is, $\Lambda(x||y) \geq 0$ if and only if $x \xrightarrow{\text{noisy}} y$.*

Proof. By Proposition 36, $\Lambda(x||y) \geq 0$ implies that $x \xrightarrow{\text{noisy}} y \otimes s_I$ with $I \geq 0$ and because marginalizing over s_I is a free operation, this implies that $x \xrightarrow{\text{noisy}} y$. Similarly, by Proposition 37, $\Lambda(x||y) < 0$ implies that $x \not\xrightarrow{\text{noisy}} y$ by noisy operations. ■

This criterion was first proposed as a necessary and sufficient condition for state conversion in (Egloff *et al.*, 2012). In fact, they provided the generalization of this condition for the resource theory of athermality. The criterion was rederived in (Faist *et al.*, 2012).

D. Catalysis

Definition 40 *In any given resource theory, if $x \not\xrightarrow{\text{noisy}} y$ under the free operations, but there exists a state z such that $x \otimes z \mapsto y \otimes z$, then z is said to be a catalyst for the conversion of x to y .*

In the resource theory of nonuniformity, there are non-trivial examples of catalysis. This is true both for state conversions between states of equal dimension and between states of unequal dimension. In the latter case, we have the following example, adapted from (Jonathan and Plenio, 1999). The two states are

$$x = (0.5, 0.25, 0.25, 0), \quad (5.16)$$

$$y = (0.8, 0.2), \quad (5.17)$$

and the catalyst is

$$z = (0.6, 0.4). \quad (5.18)$$

Here, $x \not\xrightarrow{\text{noisy}} y$ by noisy operations, but $x \otimes z \xrightarrow{\text{noisy}} y \otimes z$. Specifically, there is a permutation taking $x \otimes z$ to $y \otimes m^{(2)} \otimes z$ where $m^{(2)}$ is the uniform state on a bit. One then marginalizes over $m^{(2)}$ to obtain $y \otimes z$.

Although we have seen that Lorenz curves are extremely useful tools for many notions of state conversion under noisy operations, they are of limited usefulness in understanding catalysis. Nonetheless, they do make evident one simple result:

Proposition 41 *A sharp state is useless as a catalyst, that is, for all finite $I \in \mathbb{R}_+$, $x \otimes s_I \xrightarrow{\text{noisy}} y \otimes s_I$ if and only if $x \xrightarrow{\text{noisy}} y$.*¹⁹

To see that this is the case, recall from Lemma 33 that the operation of joining a sharp state with nonuniformity I to a state x , $x \mapsto x \otimes s_I$, is represented in terms of Lorenz curves as a 2^I -fold compression along the u -axis towards the origin, and appending a tail of the appropriate length. As such, if the Lorenz curve of x does not lie everywhere above or on the Lorenz curve of y , then the same relationship will hold between the Lorenz curve of $x \otimes s_I$ and that of $y \otimes s_I$.

As an aside, this result has an interesting consequence for the role of ideal measurements in the resource theory of nonuniformity. It is clear that the capacity to implement an ideal measurement requires a resource of nonuniformity because it requires the pointer to be initialized in a sharp state. Nonetheless, one might wonder whether such an ideal measurement device might sometimes be used in order to make possible a state conversion that would otherwise be impossible, *without* consuming any of the nonuniformity in the device. In short, one might wonder whether an ideal measurement device might sometimes act as a catalyst. However, insofar as such a device is modelled by a sharp state, Proposition 41 implies that it is useless as a catalyst. Whatever state conversion can be done with a measurement device that is used catalytically can also be achieved without it.

Proposition 41 also implies that removing zero components from a state (or from another state that is noisy-equivalent to it) cannot change its effectiveness as a catalyst. To express this claim compactly, it is useful to introduce a notation for removing zero components.

Definition 42 *For any state x and $I \leq I_0(x)$, define $x^{(I)}$ to be the smallest-dimensional state such that $x^{(I)} \otimes s_I$ is noisy-equivalent to x .*

Clearly, the Lorenz curve of $x^{(I)}$ is that of x stretched by a factor of 2^I only the u -axis, thereby shortening its tail length. If $I = I_0(x)$, then the Lorenz curve of $x^{(I)}$ is that of x stretched to the point where its tail length goes to zero. The map $x \mapsto x^{(I)}$ will be referred to as a *truncation* of x .

Corollary 43 *For any $I \leq I_0(z)$, z catalyzes the conversion of x to y under noisy operations if and only if $z^{(I)}$ does as well, that is, $x \otimes z \xrightarrow{\text{noisy}} y \otimes z$ if and only if $x \otimes z^{(I)} \xrightarrow{\text{noisy}} y \otimes z^{(I)}$.*

To see that this is the case, it suffices to note that z is noisy-equivalent to $z^{(I)} \otimes s_I$, so that $x \otimes z \xrightarrow{\text{noisy}} y \otimes z$ if

and only if $x \otimes z^{(I)} \otimes s_I \xrightarrow{\text{noisy}} y \otimes z^{(I)} \otimes s_I$, and then to apply Proposition 41. It follows that states which act as nontrivial catalysts do not need to contain any zeros.

1. Quasi-order of states under noisy operations assisted by a catalyst

A quasi-order over states that is induced by a set of free operations assisted by a catalyst has been called a *trumping* quasi-order (Nielsen, 2002). We will refer to the trumping quasi-order induced by noisy operations as the *noisy-trumping* quasi-order.

At first glance, one might expect that the definition of noisy-trumping ought to be that x noisy-trumps y if there exists a catalyst z such that $x \otimes z \xrightarrow{\text{noisy}} y \otimes z$. However, we will see that this definition leads to some mathematical complications that are not physically significant, and we will consequently be led to consider modifications of this definition wherein either the input or the output state is allowed to be arbitrarily well approximated by x or y respectively. This move is analogous to how we chose to define noisy operations (Definition 1) as those that can be arbitrarily well approximated by maps of the form (2.1) rather than those that are precisely of that form.

We start by recapitulating the part of the results of (Klimesh, 2007) and (Turgut, 2007) which is the basis of all that follows.

Lemma 44 (Klimesh (2007)) *Let x and y be states of the same dimensionality (i.e. $d_x = d_y =: d$) that do not both contain zero components and that satisfy $x^\downarrow \neq y^\downarrow$. Then there is a catalyst z such that $x \otimes z \xrightarrow{\text{noisy}} y \otimes z$ if and only if $f_r(x) > f_r(y)$ for all $r \in \mathbb{R}$, where*

$$f_r(x) = \begin{cases} \ln \sum_{i=1}^d x_i^r & \text{if } r > 1, \\ \sum_{i=1}^d x_i \ln x_i & \text{if } r = 1, \\ -\ln \sum_{i=1}^d x_i^r & \text{if } 0 < r < 1, \\ -\sum_{i=1}^d \ln x_i & \text{if } r = 0, \\ \ln \sum_{i=1}^d x_i^r & \text{if } r < 0. \end{cases}$$

This result resembles our notion of a *complete set of monotones*, introduced in Subsection IV.A.3 as describing a set of monotones that completely characterize state transformations: the lemma states that the functions f_r almost have this property – but not exactly. Namely, while $f_r(x) > f_r(y)$ for all r implies that x can be catalytically converted to y , we cannot infer anything in the case where $f_r(x) \geq f_r(y)$ (and equality, for example, is attained at only one single value of r).

Not only is this at odds with the definition of a complete set of monotones; it also indicates a certain unphysical kind of discontinuity. Imagine some fixed state x , and a sequence of states y_n , all of the same dimensionality, which converge to y in the limit of large n , i.e. $\lim_{n \rightarrow \infty} y_n = y$. It may well be the case that $f_r(x) > f_r(y_n)$ for all $r \in \mathbb{R}$ and $n \in \mathbb{N}$, but that only the non-strict inequality survives the limit, i.e. $f_r(x) \geq f_r(y)$

¹⁹ This result is the analogue of the fact that maximally entangled states are useless as catalysts in entanglement theory (Jonathan and Plenio, 1999).

with actual equality for some r . In this case, we could convert x catalytically into *any* of the states y_n , but *not* into the state y .

The physical interpretation would be odd: it would tell us that we cannot produce y perfectly (catalytically from x), but we can produce it to arbitrary accuracy. However, in actual physical situations, we can *never* expect to produce any state perfectly; conversion to arbitrary accuracy is the best we can hope for. Therefore, we would like to have a definition of trumping that takes this physical limitation into account, and that includes situations where one can obtain y from x catalytically to arbitrary accuracy, even if one cannot obtain it perfectly.

The most straightforward way to obtain a definition like this would be to say that x noisy-trumps y if and only if x can be catalytically converted to an arbitrarily good approximation of y by a suitable catalyst. In Appendix B, we will analyze this definition, recovering and elaborating some results from (Brandão *et al.*, 2015). However, we will see that this definition still contains a discontinuity that has no physical counterpart, namely, that input states which contain zero components behave differently from input states that have arbitrarily small components, even though the two could never be physically distinguished.

To achieve a physically sensible definition of noisy trumping, we must instead allow that the input state required to form y catalytically be merely arbitrarily close to the noisy equivalence class of x , rather than a perfect copy of x . In order to avoid the mathematical subtleties of defining a metric over a set of states of differing dimensions, we will adopt a specific type of approximation to the noisy equivalence class of x , namely, the state that is the composite of x and a sharp state that is arbitrarily close to the uniform state. This sort of definition was also suggested in (Brandão *et al.*, 2015) where it was described as allowing some nonuniformity to be consumed in the process, as long as the amount of nonuniformity can be made arbitrarily small.

Definition 45 (Noisy trumping) *We say that x noisy-trumps y if for any $\delta > 0$, no matter how small, there is a catalyst z such that*

$$x \otimes s_\delta \otimes z \xrightarrow{\text{noisy}} y \otimes z.$$

It turns out that the ability to consume an arbitrarily small amount of nonuniformity unlocks potential state transformations that are otherwise impossible.

Clearly, if $x \otimes z \xrightarrow{\text{noisy}} y \otimes z$ for some catalyst z , then x noisy-trumps y (because for the input state $x \otimes s_\delta \otimes z$, one can simply marginalize over s_δ and then implement $x \otimes z \xrightarrow{\text{noisy}} y \otimes z$). However, the converse is not true: it is impossible to obtain $x \otimes z \xrightarrow{\text{noisy}} y \otimes z$ as a formal “limit $\delta \rightarrow 0$ ” of the definition of noisy trumping. This will be established below, by showing that the conditions for noisy-trumping are different from the conditions in Lemma 44.

It is no restriction to demand that the target state y be produced perfectly: as we show in Corollary 47 below, it is an automatic consequence of Definition 45 that the ability to produce arbitrarily good approximations of y implies the ability to produce y perfectly.

Note also that Definition 45 requires that the final state of the catalyst be *precisely* equal to its initial state. This implies that the catalyst can be reused arbitrarily many times and consequently that any nonuniformity required to form y must have been drawn from x , or a state arbitrarily close to it, and not from the catalyst.²⁰

Finally, note that in this article, we refer to state conversions between arbitrarily good approximations of states as *exact*. A state conversion is only described as *approximate* if the degree of approximation is finite. Section VI considers approximate state conversion.

The results by Klimesh and Turgut, Lemma 44, imply the following characterization of noisy-trumping:

Lemma 46 (Conditions for noisy-trumping) *Let x and y be any pair of states. Then x noisy-trumps y if and only if*

$$I_p(x) \geq I_p(y) \text{ for all } p \geq 0. \quad (5.19)$$

In other words, the set of Rényi p -nonuniformities for nonnegative p , $\{I_p : p \geq 0\}$ defined in Table I (the $p = 1$ case corresponds to the Shannon nonuniformity), is a complete set of monotones for the noisy trumping relation.

Proof. Necessity of these conditions for noisy-trumping is straightforward to prove. Applying the fact that every I_p is a nonuniformity monotone (cf. Subsection IV.A.5) to Definition 45, we obtain

$$I_p(x \otimes s_\delta \otimes z) \geq I_p(y \otimes z)$$

for every $\delta > 0$, and using additivity of I_p as well as $I_p(s_\delta) = \delta$, this becomes

$$I_p(x) + \delta \geq I_p(y) \quad \text{for all } \delta > 0,$$

which implies that $I_p(x) \geq I_p(y)$.

Sufficiency of the conditions for noisy-trumping are much more difficult to prove. We will do so by appealing to the results of (Klimesh, 2007), summarized in Lemma 44.

We begin by noting that if the states under consideration are of unequal dimensions, $d_x \neq d_y$, then we can simply replace x by $x \otimes m^{(d_y)}$ and replace y by $y \otimes m^{(d_x)}$,

²⁰ Note that this verdict on the reusability of the catalyst would not change even if we interpreted the sharp state s_δ as being a modification of the initial state of the catalyst rather than a modification of the initial state of x . For any given number N of reuses of the catalyst, one could still take $N\delta$ to be arbitrarily small.

because the replacement does not alter the values of any of the I_p 's, the property of containing zeros or not, or the noisy-trumping property. Thus, we may assume that $d_x = d_y$.

So suppose that $I_p(x) \geq I_p(y)$ for all $p \geq 0$. Fix any $\delta > 0$ which is the logarithm of a rational number. We start by removing common zeros from x and y in the sense of Definition 42. Set $y' := y^{(I_0(y))}$. Since $I_0(x) \geq I_0(y)$, the state $x' := x^{(I_0(y))}$ is well-defined. We will be considering a state conversion with initial state $x' \otimes s_\delta$ and a final state in the noisy equivalence class of y' . To apply Lemma 44, we need to ensure that the dimensions of the initial and final states are equal, so we take the final state to be $y' \otimes m^{(d_{s_\delta})}$ where d_{s_δ} is the dimension of the space on which s_δ is defined. We now show that $I_p(x' \otimes s_\delta) > I_p(y' \otimes m^{(d_{s_\delta})})$ for all $p \geq 0$.

By definition of x' , $x' \otimes s_{I_0(y)}$ is noisy-equivalent to x . Recalling the additivity of the I_p 's and the fact that $I_p(s_I) = I$, we have that $I_p(x') + I_0(y) = I_p(x)$. For similar reasons, $I_p(y') + I_0(y) = I_p(y)$. This implies that

$$\begin{aligned} I_p(x' \otimes s_\delta) &= I_p(x) - I_0(y) + \delta \\ &> I_p(y) - I_0(y) = I_p(y' \otimes m^{(d_{s_\delta})}) \end{aligned}$$

for all $p > 0$. We note that the monotonicity of the family of functions $\{f_p \mid p > 0\}$ is equivalent to the monotonicity of the family $\{I_p \mid p > 0\}$ of Rényi p -nonuniformities. Therefore, we obtain $f_p(x' \otimes s_\delta) > f_p(y' \otimes m^{(d_{s_\delta})})$ for all $p > 0$. Since $x' \otimes s_\delta$ contains zeros, but $y' \otimes m^{(d_{s_\delta})}$ does not, it follows that for all $p \leq 0$ we have $\infty = f_p(x' \otimes s_\delta) > f_p(y' \otimes m^{(d_{s_\delta})})$. Thus, Lemma 44 shows that there is a catalyst z such that

$$x' \otimes s_\delta \otimes z \xrightarrow{\text{noisy}} y' \otimes m^{(d_{s_\delta})} \otimes z.$$

By adding on a sharp state $s_{I_0(y)}$ on both sides (which is not touched during the state conversion), and using again the noisy-equivalence of $x' \otimes s_{I_0(y)}$ and x (and similarly for y), and the fact that one can marginalize over the uniform state, we see that $x \otimes s_\delta \otimes z \xrightarrow{\text{noisy}} y \otimes z$. Since $\delta > 0$ was chosen arbitrarily, this proves that x noisy-trumps y . ■

An immediate consequence of Lemma 46 is a certain continuity (or closedness) property which is desirable from the point of view of physical interpretation.

Corollary 47 *Suppose that x is any state, and $(y_n)_{n \in \mathbb{N}}$ is any sequence of states of common dimensionality which converges to some state y , i.e. $\lim_{n \rightarrow \infty} y_n = y$. If x noisy-trumps every y_n then x noisy-trumps y .*

Proof. According to Lemma 46, if x noisy-trumps every y_n then $I_p(x) \geq I_p(y_n)$ for every $n \in \mathbb{N}$ and $p \geq 0$. Since every I_p is continuous if $p > 0$, it follows that $I_p(x) \geq I_p(y)$ for $p > 0$. Even though I_0 is not continuous, we can express it as $I_0 = \lim_{p \searrow 0} I_p$, and so we also obtain $I_0(x) \geq I_0(y)$ by taking the limit $p \searrow 0$.

Invoking Lemma 46 again shows that x noisy-trumps y . ■

Lemma 46 can also be expressed in terms of the existence of a complete witness for noisy-trumping, as follows.

Definition 48 *For an arbitrary pair of states x and y , define*

$$\Lambda_{\text{cat}}(x \| y) := \inf_{p \geq 0} (I_p(x) - I_p(y)).$$

Proposition 49 *For an arbitrary pair of states x and y , x noisy-trumps y if and only if*

$$\Lambda_{\text{cat}}(x \| y) \geq 0. \quad (5.20)$$

Equivalently, the function $\Lambda_{\text{cat}}(x \| y)$ is a complete witness for noisy-trumping.

The necessary and sufficient conditions for noisy-trumping are a strict subset of those for noisy-majorization because noisy-majorization implies noisy-trumping but not vice-versa. It follows that one must add conditions to those of Lemma 46 in order to characterize noisy-majorization. In other words, Rényi p -nonuniformities alone cannot decide whether a given non-catalytic state conversion is possible. This is true even if we take the Rényi p -nonuniformities of orders $p < 0$ into account. The reason is as follows. Every Rényi p -nonuniformity is additive and therefore is nonincreasing under $x \xrightarrow{\text{noisy}} y$ if and only if it is nonincreasing under $x \otimes z \xrightarrow{\text{noisy}} y \otimes z$ for any state z . To do justice to the distinction between noisy-majorization and noisy-trumping, therefore, one must consider one or more monotones that are *not* additive, hence one must go beyond the Rényi nonuniformities.

2. Nonuniformity of formation and distillable nonuniformity in the presence of a catalyst

The nonuniformity of formation and the distillable nonuniformity (as determined in Proposition 30 and 31) are not changed by having access to a catalyst.

Corollary 50 *x noisy-trumps s_I if and only if*

$$I \leq I_0(x),$$

that is, the maximum nonuniformity of sharp state that can be distilled from x under noisy operations assisted by a catalyst is $I_0(x)$.

Proof. Sufficiency is trivial because if $I \leq I_0(x)$, then by the result on distillable nonuniformity (Proposition 30) it follows that $x \xrightarrow{\text{noisy}} s_I$, so that one does not even require a catalyst to achieve the conversion. It suffices, therefore, to prove necessity. But this follows from Lemma 46, which shows that I_0 is a noisy-trumping monotone, hence $I_0(x) \geq I_0(s_I) = I$. ■

Corollary 51 s_I noisy-trumps x if and only if

$$I \geq I_\infty(x),$$

that is, the minimum nonuniformity of sharp state that is required to form x under noisy operations assisted by a catalyst is $I_\infty(x)$.

Proof. Again, sufficiency is straightforward to establish because if $I \geq I_\infty(x)$ then by the result on the nonuniformity of formation (Proposition 31) we can infer that $s_I \xrightarrow{\text{noisy}} x$, so that the conversion can be achieved even without a catalyst. To establish necessity, we use again Lemma 46: the limit $p \rightarrow \infty$ shows that I_∞ is a noisy-trumping monotone, hence $I = I_\infty(s_I) \geq I_\infty(x)$. ■

3. Nonuniformity cost and yield of catalytic state conversion

In catalytic state conversion, we also have a result akin to Propositions 36 and 37. This was described in the context of the theory of athermality in (Brandão *et al.*, 2015).

Proposition 52 If $\Lambda_{\text{cat}}(x||y) > 0$, then x noisy-trumps $y \otimes s_I$ if and only if

$$I \leq \Lambda_{\text{cat}}(x||y).$$

That is, if $\Lambda_{\text{cat}}(x||y)$ is positive, then using a catalyst one can distill a sharp state of nonuniformity at most $\Lambda_{\text{cat}}(x||y)$ in addition to achieving the conversion of x to y .

Proof. We note that

$$\begin{aligned} \Lambda_{\text{cat}}(x||y \otimes s_I) &= \inf_{p \geq 0} (I_p(x) - I_p(y \otimes s_I)) \\ &= \inf_{p \geq 0} (I_p(x) - I_p(y)) - I \\ &= \Lambda_{\text{cat}}(x||y) - I, \end{aligned}$$

where we have made use of the additivity of I_p and the fact that $I_p(s_I) = I$ for $p \geq 0$. According to Proposition 49, x noisy-trumps $y \otimes s_I$ if and only if $\Lambda_{\text{cat}}(x||y \otimes s_I)$ is non-negative, hence if and only if $\Lambda_{\text{cat}}(x||y) \geq I$. ■

Proposition 53 If $\Lambda_{\text{cat}}(x||y) < 0$, then $x \otimes s_I$ noisy-trumps y if and only if

$$I \geq -\Lambda_{\text{cat}}(x||y).$$

That is, if $\Lambda_{\text{cat}}(x||y)$ is negative, then in order to achieve the conversion of x to y with the aid of a catalyst, it costs a sharp state of nonuniformity at least $|\Lambda_{\text{cat}}(x||y)|$.

Proof. We have

$$\begin{aligned} \Lambda_{\text{cat}}(x \otimes s_I||y) &= \inf_{p \geq 0} (I_p(x \otimes s_I) - I_p(y)) \\ &= I + \inf_{p \geq 0} (I_p(x) - I_p(y)) \\ &= I + \Lambda_{\text{cat}}(x||y). \end{aligned}$$

According to Proposition 49, $x \otimes s_I$ noisy-trumps y if and only if $\Lambda_{\text{cat}}(x \otimes s_I||y)$ is non-negative, hence if and only if $I \geq -\Lambda_{\text{cat}}(x||y)$. ■

We again recover the results about formation and distillation as special cases: the distillable nonuniformity of x under catalyzed noisy operations is simply the nonuniformity yield of the conversion of x to the uniform state m , and $\Lambda_{\text{cat}}(x||m) = \inf_{p \in \mathbb{R}} I_p(x) = I_0(x)$; similarly, the nonuniformity of formation of x in the presence of a catalyst is recovered from $-\Lambda_{\text{cat}}(m||x) = \sup_{p \geq 0} I_p(x) = I_\infty(x)$.

E. Inadequacy of the second law as a criterion for state conversion

In the context of the resource theory of nonuniformity, we might imagine that the role of the second law of thermodynamics is to provide a criterion under which a state conversion $x \mapsto y$ is possible by noisy operations. The usual statement of the second law is, however, inadequate to this task.

The second law asserts that the Shannon entropy must be nondecreasing in the evolution of a state. In the context of states of equal dimension, this is equivalent to saying that the Shannon negentropy or the Shannon nonuniformity must be nonincreasing.

The first inadequacy is that for conversions between states of *unequal* dimension, it is insufficient to look at the Shannon entropy (or the Shannon negentropy). One must look instead at the Shannon *nonuniformity*. For instance, for the pair of states $y = (1/2, 1/2)$ and $x = (1/3, 1/3, 1/3, 0)$, the entropies are $H(y) = 1$ and $H(x) = \log 3$, so that $H(y) < H(x)$. And yet the state conversion $x \mapsto y$ is possible by noisy operations, despite involving a *decrease* in entropy. This is to be expected, because marginalization is an instance of a noisy operation and it can easily reduce the entropy. The Shannon nonuniformity, on the other hand, is nonincreasing under noisy operations. In particular, in our example $I(x) = \log(4/3)$ while $I(y) = 0$.

The second inadequacy regards sufficiency of the condition. Even if we consider a conversion between states of equal dimension, so that the Shannon nonuniformity is nonincreasing if and only if the Shannon entropy is nondecreasing, the nondecrease of the Shannon entropy is not a sufficient condition for the state conversion to be possible. This is because one can easily construct examples wherein $H(x) \leq H(y)$ (equivalently $I(x) \geq I(y)$) but it is not the case that $x \xrightarrow{\text{noisy}} y$. For instance, for states $x = (2/3, 1/6, 1/6)$ and $y = (1/2, 1/2, 0)$ one can easily verify (for instance, via the Lorenz curves) that $x \not\xrightarrow{\text{noisy}} y$ by noisy operations. Yet, $H(x) \simeq 0.811$ and $H(y) = 1$.

Furthermore, one can never rehabilitate the second law by simply replacing entropy with some other nonuniformity monotone because the set of states forms a quasi-order under noisy operations, and therefore no single nonuniformity monotone can decide every state conver-

sion problem. To provide necessary and sufficient conditions for state conversion, one must look at the values of a *complete set* of nonuniformity monotones. To our knowledge, this point was first made in (Ruch and Lesche, 1978).

Note, however, that if a complete set of monotones is of infinite cardinality, then such a criterion will not be of much practical use. Thus, we would like to also require that our criterion for state conversion be decidable by finite means. In many cases this can be accomplished by considering *state conversion witnesses*, defined in Section II.A.

In this article, we have seen two examples of complete witnesses for state conversion under noisy operations. These were defined in Eqs. (3.12) and (5.12). It follows that a rehabilitated second law of thermodynamics, in the context of the resource theory of nonuniformity, can be stated thus: $x \xrightarrow{\text{noisy}} y$ if and only if either of the following equivalent conditions hold:

- $\Delta(x||y) \geq 0$ where

$$\Delta(x||y) := \min_{k \in \{1, \dots, d_y\}} (L_x(k/d_x) - L_y(k/d_y))$$

- $\Lambda(x||y) \geq 0$ where

$$\Lambda(x||y) := \max \{ \lambda : L_x(u) \geq L_y(2^\lambda u) \quad \forall u \in \mathcal{U} \},$$

with

$$\mathcal{U} := \left(\left\{ \frac{1}{d_x}, \frac{2}{d_x}, \dots, 1 \right\} \cup \left\{ \frac{1}{d_y}, \frac{2}{d_y}, \dots, 1 \right\} \right) \cap (0, u_*]$$

$$\text{and } u_* := \min\{1, 2^{-\lambda}\}.$$

Evaluating these functions requires an optimization over a finite number of expressions, so that the conversion problem becomes decidable by finite means.

Note that in the resource theory of athermality, the free energy plays the role of the Shannon nonuniformity, but because a single athermality monotone cannot capture the quasi-order of athermal states, the free energy is also inadequate to the task of providing necessary and sufficient conditions for state conversion.

One might think that the role of the second law is to provide a criterion for *catalytic* state conversion rather than the regular variety. Indeed, many statements of the second law allow for the environment of the system to undergo a change, as long as the process is cyclic, which is to say that in the end, the environment is returned to its initial state. In the language of resource theories, the environment is allowed to be a catalyst for the state transition (Brandão *et al.*, 2015). However, even under this interpretation, the standard formulation of the second law is inadequate. For states of equal dimension, the nondecrease of entropy remains only a necessary, but not a sufficient, condition for the state conversion to be possible. The quasi-order of states under catalytic noisy operations cannot be captured by any single nonuniformity monotone.

However, a complete witness for state conversion under catalytic noisy operations is implied by Propositions 52 and 53, namely: x noisy-trumps y if and only if

- $\Lambda_{\text{cat}}(x||y) \geq 0$ where

$$\Lambda_{\text{cat}}(x||y) := \inf_{p \geq 0} (I_p(x) - I_p(y)).$$

It remains an open problem to provide a criterion for catalytic state conversion that is decidable by finite means.

VI. APPROXIMATE STATE CONVERSION

A. Metrics over state space

So far, we have discussed the question under what conditions noisy operations can transform a given state into another state *exactly* or to arbitrary precision. However, in most physically relevant cases, this is too stringent a requirement, and a different problem becomes important: *Under what conditions can noisy operations transform a given state into another state which is ε -close to the desired target state for some fixed ε ?*

Two immediate questions arise in the study of this problem. First, what kind of distance measure on the set of states should we use? Second, can the quantum problem be reduced to the classical problem in the same way as in the case of exact state conversion?

The answer to the latter question turns out to be yes, given that we restrict our attention to distance measures—that is, metrics—on the quantum states and the classical probability distributions that have the crucial property of *contractivity*:

Definition 54 *A contractive metric on the quantum states is a map \mathcal{D} that assigns to two quantum states ρ, σ on the same underlying Hilbert space of dimension $d_\rho = d_\sigma$ a non-negative real number $\mathcal{D}(\rho, \sigma)$ satisfying the axioms of a metric*

- $\mathcal{D}(\rho, \sigma) = \mathcal{D}(\sigma, \rho)$,
- $\mathcal{D}(\rho, \sigma) = 0 \Leftrightarrow \rho = \sigma$,
- $\mathcal{D}(\rho, \sigma) \leq \mathcal{D}(\rho, \tau) + \mathcal{D}(\tau, \sigma)$ for all quantum states τ with $d_\tau = d_\sigma = d_\rho$,

such that every completely positive, trace-preserving map Φ is a contraction, i.e.

$$\mathcal{D}(\Phi(\rho), \Phi(\sigma)) \leq \mathcal{D}(\rho, \sigma).$$

Analogously, a contractive metric on the classical (discrete) probability distributions is a map \mathcal{D} that assigns to two probability distributions x, y with the same dimension of the underlying sample space $d_x = d_y$ a non-negative real number $\mathcal{D}(x, y)$ that satisfies the classical analogs of the axioms above.

In this definition, the classical analog of a completely positive trace-preserving map Φ is a channel Φ , i.e. a linear map that is described by a finite stochastic matrix, or equivalently, a linear map that maps probability distributions to probability distributions.

Note that $\mathcal{D}(\rho, \sigma)$, respectively $\mathcal{D}(x, y)$, is undefined if $d_\rho \neq d_\sigma$, respectively $d_x \neq d_y$; however, this does not mean that the values of \mathcal{D} for different dimensions are completely unrelated. In fact, contractivity implies some relations between the distances of states of different dimensionalities. For example, if τ is an arbitrary fixed finite-dimensional quantum state, then the map $\rho \mapsto \rho \otimes \tau$ is a quantum operation, hence $\mathcal{D}(\rho \otimes \tau, \sigma \otimes \tau) \leq \mathcal{D}(\rho, \sigma)$. On the other hand, taking the partial trace over the second system is a quantum operation too, implying the converse inequality, such that we get

$$\mathcal{D}(\rho \otimes \tau, \sigma \otimes \tau) = \mathcal{D}(\rho, \sigma).$$

If we are given a contractive metric \mathcal{D} on the quantum states, then we can construct a contractive metric on the classical probability distributions by restricting \mathcal{D} to diagonal matrices; we will call this the classical metric that *corresponds to* the respective quantum metric. For example, the trace distance on quantum states

$$\mathcal{D}_{\text{tr}}(\rho, \sigma) := \frac{1}{2} \text{tr}|\rho - \sigma| = \frac{1}{2} \|\rho - \sigma\|_1$$

has the corresponding classical metric

$$\mathcal{D}_{\text{tr}}(x, y) = \frac{1}{2} \sum_{i=1}^{d_x} |x_i - y_i| \quad (6.1)$$

(note that $d_x = d_y$, otherwise the expression $\mathcal{D}(x, y)$ is undefined). Another contractive metric that satisfies the conditions of our definition is the *purified distance* (Tomamichel, 2012)

$$\mathcal{D}_p(\rho, \sigma) := \sqrt{1 - F(\rho, \sigma)^2}$$

with $F(\rho, \sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1$ the fidelity, which has the corresponding classical contractive metric

$$\mathcal{D}_p(x, y) = \sqrt{1 - \left(\sum_i \sqrt{x_i y_i} \right)^2}. \quad (6.2)$$

A consequence of contractivity is unitary equivalence: if U is unitary, then

$$\mathcal{D}(U\rho U^\dagger, U\sigma U^\dagger) = \mathcal{D}(\rho, \sigma).$$

This is because the unitary map as well as its inverse are quantum operations, and therefore contractivity implies inequalities in both directions. It follows that classical contractive metrics are left invariant if we permute the entries of the two probability vectors in its argument in the same way.

In the case of contractive metrics, approximate state transformation in the quantum case can be reduced to the classical problem:

Lemma 55 *Let $\varepsilon > 0$, let ρ and σ be two quantum states, and let \mathcal{D} be any contractive metric on the quantum states. Then the following two statements are equivalent:*

- (i) *There exists a quantum state $\tilde{\sigma}$ with $\mathcal{D}(\sigma, \tilde{\sigma}) \leq \varepsilon$ such that noisy quantum operations can transform ρ to $\tilde{\sigma}$.*
- (ii) *There exists a distribution \tilde{s} with $\mathcal{D}(\lambda(\sigma), \tilde{s}) \leq \varepsilon$ such that noisy classical operations can transform $\lambda(\rho)$ to \tilde{s} .*

Here, the metric \mathcal{D} in (ii) is the classical contractive metric corresponding to the quantum contractive metric \mathcal{D} . Note that in general $\tilde{s} \neq \lambda(\tilde{\sigma})$.

Proof. (ii) \Rightarrow (i): with $\sigma = \sum_k \lambda_k(\sigma) |k\rangle\langle k|$, define $\tilde{\sigma} := \sum_k \tilde{s}_k |k\rangle\langle k|$, then $\mathcal{D}(\sigma, \tilde{\sigma}) \leq \varepsilon$. Since noisy classical operations can transform $\lambda(\rho)$ into $\lambda(\tilde{\sigma})$, Lemma 7 applies and proves that noisy operations can transform ρ into $\tilde{\sigma}$.

To see (i) \Rightarrow (ii), let Φ be the quantum operation that dephases in the eigenbasis of σ ; i.e. $\Phi(\tau) := \sum_k \langle k|\tau|k\rangle |k\rangle\langle k|$, and set $\tilde{s}_k := \langle k|\tilde{\sigma}|k\rangle$, then $\Phi(\tilde{\sigma}) = \sum_k \tilde{s}_k |k\rangle\langle k|$, and

$$\mathcal{D}(\lambda(\sigma), \tilde{s}) = \mathcal{D}(\sigma, \Phi(\tilde{\sigma})) = \mathcal{D}(\Phi(\sigma), \Phi(\tilde{\sigma})) \leq \mathcal{D}(\sigma, \tilde{\sigma}) \leq \varepsilon.$$

Again, due to Lemma 7, the assumption that $\rho \xrightarrow{\text{noisy}} \tilde{\sigma}$ implies that $\lambda(\rho) \xrightarrow{\text{noisy}} \lambda(\tilde{\sigma})$. Furthermore, according to Exercise II.1.12, ‘‘Schur’s Theorem’’ in (Bhatia, 1997), the eigenvalues of $\tilde{\sigma}$ majorize the diagonal elements of $\tilde{\sigma}$ in the eigenbasis of σ , hence $\lambda(\tilde{\sigma}) \succ \tilde{s}$, and consequently $\lambda(\tilde{\sigma}) \xrightarrow{\text{noisy}} \tilde{s}$. It follows that $\lambda(\rho) \xrightarrow{\text{noisy}} \tilde{s}$. ■

This lemma allows us to study approximate state conversion in purely classical terms. Hence, in the following, when we refer to a contractive metric \mathcal{D} , we always assume that it is a contractive metric on the *classical probability distributions*.

For convenience, we will make use of the following notational convention:

Definition 56 (ε -approximate state conversion)

We write $x \xrightarrow{\varepsilon\text{-noisy}} y$ if there exists a noisy classical operation taking x to a state that is ε -close to y relative to a contractive metric \mathcal{D} .

The following transitivity property will be used several times: if we have both

$$x \xrightarrow{\varepsilon\text{-noisy}} y \quad \text{and} \quad y \xrightarrow{\delta\text{-noisy}} z$$

then it follows that

$$x \xrightarrow{(\varepsilon + \delta)\text{-noisy}} z.$$

B. Smoothed entropies

As we have seen in Subsection V.B, the order- ∞ and order-0 Rényi nonuniformities determine the single-shot nonuniformity of formation and distillable nonuniformity of a state. When the state preparation is allowed to be approximate, these quantities turn out to be given by smoothed versions of these Rényi nonuniformities. We begin, therefore, with a discussion of smoothing.

Definition 57 Fix any contractive metric \mathcal{D} . For probability distributions x and $\varepsilon \geq 0$, define the smooth order-0 and order- ∞ Rényi entropies as

$$\begin{aligned} H_\infty^\varepsilon(x) &:= \max_{x': \mathcal{D}(x, x') \leq \varepsilon} H_\infty(x'), \\ H_0^\varepsilon(x) &:= \min_{x': \mathcal{D}(x, x') \leq \varepsilon} H_0(x'). \end{aligned}$$

Similarly, we define the smooth order-0 and order- ∞ Rényi nonuniformities as

$$\begin{aligned} I_\infty^\varepsilon(x) &:= \log d_x - H_\infty^\varepsilon(x) = \min_{x': \mathcal{D}(x, x') \leq \varepsilon} I_\infty(x'), \\ I_0^\varepsilon(x) &:= \log d_x - H_0^\varepsilon(x) = \max_{x': \mathcal{D}(x, x') \leq \varepsilon} I_0(x'). \end{aligned}$$

It is clear that the values of these quantities depend on the choice of metric \mathcal{D} . In the following, we will always assume that the reader is aware of this freedom of choice, and we will not always repeat that the statements are contingent on the choice of \mathcal{D} .

Note that the optimizations in Definition 57 are over *normalized* classical probability distributions x' . This differs from some places in the literature where subnormalized (quantum) states are taken into account, for example in (Tomamichel, 2012).

In Lemma 55 above, we have shown that we can without loss of generality restrict our analysis to classical probability distributions. However, some readers may be interested in translating some results of this section directly into the quantum case, in order to apply them in other contexts. This turns out to be possible whenever we speak about the smooth Rényi ∞ -entropy: defining the quantum version of the smooth Rényi ∞ -entropy in the usual way, the resulting quantity agrees with the classical version applied to the spectrum:

$$S_\infty^\varepsilon(\rho) := \max_{\rho': \mathcal{D}(\rho, \rho') \leq \varepsilon} S_\infty(\rho') = H_\infty^\varepsilon(\lambda(\rho)), \quad (6.3)$$

where $S_\infty(\sigma) := -\log \|\sigma\|_\infty$. [This definition resembles that in (Renner, 2008) and (Tomamichel, 2012) but involves any contractive metric and involves smoothing over only normalized states.] However, we do not know whether the analogous relation holds for the smooth Rényi 0-entropy H_0^ε in general; only that

$$S_0^\varepsilon(\rho) := \min_{\rho': \mathcal{D}(\rho, \rho') \leq \varepsilon} S_0(\rho') \leq H_0^\varepsilon(\lambda(\rho)).$$

Note that, unlike eq. (6.3), we have an *inequality* here. We do get equality in the special case where the contractive metric is chosen to be the trace distance \mathcal{D}_{tr}

from (6.1), as the following lemma asserts (proof in Appendix A):

Lemma 58 Let \mathcal{D} be any contractive metric on the quantum states for which there exists a norm $\|\cdot\|$ on the self-adjoint matrices such that $\mathcal{D}(\rho, \sigma) = \|\rho - \sigma\|$ (for example the trace distance, $\mathcal{D} = \mathcal{D}_{\text{tr}}$ with $\|\cdot\| = \frac{1}{2}\|\cdot\|_1$). Then the quantum version of the smooth max-entropy,

$$S_0^\varepsilon(\rho) := \min_{\bar{\rho}: \mathcal{D}(\rho, \bar{\rho}) \leq \varepsilon} S_0(\bar{\rho}),$$

with $S_0(\bar{\rho}) := \log \text{rank}(\bar{\rho})$, agrees with the classical smooth max-entropy on the spectrum of ρ :

$$S_0^\varepsilon(\rho) = H_0^\varepsilon(\lambda(\rho)).$$

The fact that this restriction is necessary is a first manifestation of the fact that H_0^ε behaves in some ways “less nicely” than H_∞^ε , as we will see several times below, for example in the following lemma:

Lemma 59 For every $\varepsilon > 0$, the quantity I_∞^ε is a nonuniformity monotone, but I_0^ε is not, in general.

Proof. To prove that I_∞^ε is a nonuniformity monotone, suppose that x and y are distributions such that $y = N(x)$ for some noisy operation N . Then, due to contractivity, the set of states y' with $\mathcal{D}(y, y') \leq \varepsilon$ contains the set of distributions $N(x')$ with $\mathcal{D}(x, x') \leq \varepsilon$. Moreover, we know that I_∞ (without the ε) is a nonuniformity monotone. Thus

$$\begin{aligned} I_\infty^\varepsilon(N(x)) &= \min_{y': \mathcal{D}(y, y') \leq \varepsilon} I_\infty(y') \\ &\leq \min_{x': \mathcal{D}(x, x') \leq \varepsilon} I_\infty(N(x')) \\ &\leq \min_{x': \mathcal{D}(x, x') \leq \varepsilon} I_\infty(x') = I_\infty^\varepsilon(x). \end{aligned}$$

Now we consider I_0^ε . Any nonuniformity monotone M must satisfy $M(x \otimes m^{(d)}) = M(x)$ for $m^{(d)} = (1/d, \dots, 1/d)$ the uniform state on \mathbb{R}^d (as mentioned in Eq. (4.3)). But this equation is not generally true for I_0^ε .

As a concrete counterexample, let $0 < \varepsilon < \frac{1}{2}$, $d \geq \max\{1/(2\varepsilon), 3\}$, and consider the trace distance \mathcal{D}_{tr} . If $x = m^{(2)}$ is the maximally mixed state on one bit, then $I_0^\varepsilon(x) = 0$. Let y be the distribution $(\frac{1}{d}, \frac{1}{2d}, \frac{1}{2d}, \dots, \frac{1}{2d}, 0) \in \mathbb{R}^{2d}$. Then $\mathcal{D}_{\text{tr}}(y, x \otimes m^{(d)}) = 1/(2d) \leq \varepsilon$, so $H_0^\varepsilon(x \otimes m^{(d)}) \leq H_0(y) = \log(2d - 1)$, and $I_0^\varepsilon(x \otimes m^{(d)}) \geq \log(2d) - \log(2d - 1) > 0$. ■

In the case that the reference metric is chosen to be the trace distance \mathcal{D}_{tr} , the smooth Rényi 0-entropy H_0^ε and the corresponding nonuniformity I_0^ε can be evaluated explicitly.

Lemma 60 If the trace distance \mathcal{D}_{tr} is chosen as reference metric, then any distribution x has smooth Rényi 0-entropy

$$H_0^\varepsilon(x) = \log k,$$

where $k \geq 1$ is chosen as the smallest integer such that $\sum_{i=1}^k x_i^\downarrow \geq 1 - \varepsilon$, with $x_1^\downarrow \geq x_2^\downarrow \geq \dots$ denoting the components of x in non-increasing order.

The proof is given in Appendix A.

This lemma shows how $I_0^\varepsilon(x)$ can be determined graphically in terms of the Lorenz curve of x , if the reference metric is the trace distance: denoting the two coordinates for the plot of the Lorenz curve by u and v , draw the line $v = 1 - \varepsilon$ parallel to the u -axis, and determine the left-most elbow u_0 of the Lorenz curve that lies on or above that line. Then $I_0^\varepsilon(x) = -\log u_0$. This procedure is depicted in Fig. 8.

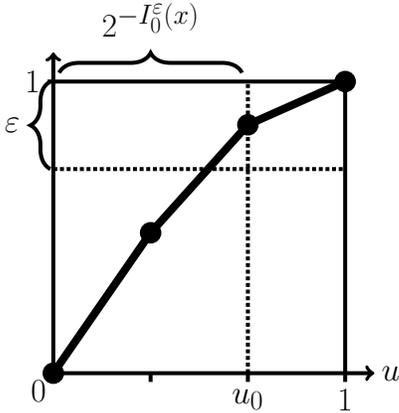


FIG. 8 If the trace distance is used as the reference metric, the smooth order-0 Rényi nonuniformity $I_0^\varepsilon(x)$ can be determined graphically in terms of the Lorenz curve of x : draw the horizontal line $v = 1 - \varepsilon$, and mark the left-most elbow that is on or above that line. Denote its u -coordinate by u_0 , then $I_0^\varepsilon(x) = -\log u_0$. Since this quantity does not only depend on the Lorenz curve, but also on the number and location of its elbows, it is clear that I_0^ε cannot be a nonuniformity monotone.

This gives a simple geometric explanation why I_0^ε is not a nonuniformity monotone, as demonstrated in Lemma 59: this quantity does not only depend on the Lorenz curve of the state, but also on its elbows. The map $x \mapsto x \otimes m^{(d)}$ changes the set of elbows (but not the Lorenz curve), and thus changes the value of I_0^ε (but not that of any nonuniformity monotone).

C. Nonuniformity of formation and distillable nonuniformity for approximate state preparation

The quantity I_∞^ε turns out to quantify the *single-shot nonuniformity of formation* if the formation is allowed to be achieved only approximately.

Lemma 61 *The minimum I such that $s_I \xrightarrow{\varepsilon\text{-noisy}} x$, is given by the ε -smoothed order- (∞) Rényi nonuniformity,*

$$I_\infty^\varepsilon(x).$$

Proof. According to Lemma 31, in order to create a state x' with $\mathcal{D}(x, x') \leq \varepsilon$, one needs a sharp state of nonuniformity at least $I_\infty(x')$. Optimizing over all possible x' yields the minimal nonuniformity of any sharp state needed to create *some* state x' which is ε -close to x :

$$\min_{x': \mathcal{D}(x, x') \leq \varepsilon} I_\infty(x') = I_\infty^\varepsilon(x).$$

■

As it turns out, it is not so simple to prove an analogous statement for the distillable nonuniformity. While one can achieve distillation of a sharp state of nonuniformity $I_0^\varepsilon(x)$ from x , this is not in general the optimal result.

Lemma 62 *It is possible to achieve $x \xrightarrow{\varepsilon\text{-noisy}} s_I$, for I equal to the ε -smoothed order-0 Rényi nonuniformity,*

$$I_0^\varepsilon(x).$$

Proof. Choose x' such that $I_0^\varepsilon(x) = I_0(x')$ and $\mathcal{D}(x, x') \leq \varepsilon$. According to Lemma 30, one can distill a sharp state s_I of nonuniformity $I = I_0(x')$ by noisy operations from x' . Denote this noisy operation by N , then $N(x') = s_I$. Set $y := N(x)$, then

$$\mathcal{D}(y, s_I) = \mathcal{D}(N(x), N(x')) \leq \mathcal{D}(x, x') \leq \varepsilon,$$

proving the claim. ■

If the expression in this lemma were optimal – that is, if $I_0^\varepsilon(x)$ gave the *maximal* nonuniformity of any sharp state that can be extracted from x – then I_0^ε would have to be a nonuniformity monotone. However, we know from Lemma 59 that it is not. Therefore, the optimal expression for distillable uniformity must be different. In fact, as a simple corollary of Lemma 62, we see immediately that nonuniformity

$$I_0^\varepsilon(x \otimes m^{(d)}) \tag{6.4}$$

for the d -dimensional uniform state $m^{(d)}$ (with d arbitrary) can be achieved as well: we can add on the uniform state $m^{(d)}$ for free, and apply Lemma 59 to $x \otimes m^{(d)}$. As shown in Lemma 59, this quantity can be larger than the corresponding one for x alone. In fact, we will show in the appendix that expression (6.4) is increasing in d , and the *maximal nonuniformity of any sharp state that can be extracted from x* is given by

$$J_0^\varepsilon(x) := \lim_{d \rightarrow \infty} I_0^\varepsilon(x \otimes m^{(d)}). \tag{6.5}$$

The result from Lemma 62, though not optimal, will be sufficient for all further applications: it will allow us to obtain conditions for state conversion and compute the optimal asymptotic transition rate. Therefore, we defer the formal proof of (6.5) to the appendix.

However, we note that in the case that the reference metric is the trace distance, the expression (6.5) can be evaluated directly in terms of the Lorenz curve of x :

adding on the uniform state $m^{(d)}$ to x with increasing d does not change the Lorenz curve, but adds more and more elbows. In the limit of $d \rightarrow \infty$, the set of elbows becomes dense on the Lorenz curve. Applying the prescription from Fig. 8 to determine $I_0^\varepsilon(x \otimes m^{(d)})$ means – in the limit of large d – to simply determine the point u_0 where the line $v = 1 - \varepsilon$ intersects the Lorenz curve. Then $J_0^\varepsilon(x) = -\log u_0$.

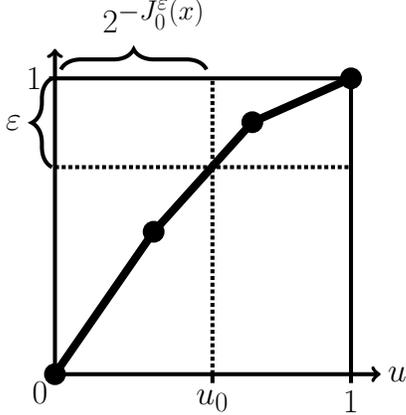


FIG. 9 If the trace distance is used as the reference metric, the distillable nonuniformity $J_0^\varepsilon(x) = \lim_{d \rightarrow \infty} I_0^\varepsilon(x \otimes m^{(d)})$ of any state x can be determined graphically by means of the Lorenz curve of x : determine the (smallest) value u_0 such that $L_x(u_0) = 1 - \varepsilon$, i.e. the intersection of the horizontal line $v = 1 - \varepsilon$ and the Lorenz curve. Then $J_0^\varepsilon(x) = -\log u_0$.

This reproduces the smoothing procedure for the “min-free energy” in (Horodecki and Oppenheim, 2013) in the case of trivial Hamiltonians $H = 0$ (cf. their Supplementary Figure S2). It is depicted in Fig. 9. It is intriguing that the Lorenz curve again turns out to be a natural and useful tool also in the case of approximate distillation of nonuniformity. However, note that the Lorenz curve procedure to determine J_0^ε is only valid if the trace distance is used as the reference metric. Expression (6.5) is more general: it is valid for all contractive metrics.

Our results above also give a qualitative confirmation of the statements in (Dahlsten *et al.*, 2009): I_0^ε quantifies the extractable nonuniformity (and thus work) in their Theorem I. Since I_∞^ε quantifies the nonuniformity needed to construct a state, any attempt to obtain more nonuniformity (and thus work) back from a given state will only be possible with very small probability of success, coming from mere chance by guessing bits correctly. This is reflected in Theorem II in their paper. However, note that the scenario in their paper differs from ours. For example, while the resource theory of nonuniformity allows one to add on uniform states for free, the number of bits (or Szilard engines) in (Dahlsten *et al.*, 2009) is fixed at some number n .

In Section V.B, a sufficient condition for exact state conversion was derived from the expressions for the nonuniformity of formation and distillable nonuniformity. In this section, we determine an analogous condition for

approximate state conversion, based on the results of the previous section. As in Subsection VI.C, we will assume that \mathcal{D} is an arbitrary contractive metric as defined in Definition 54. The two quantities I_∞^ε and I_0^ε yield a sufficient condition for approximate state interconvertibility:

Lemma 63 *If x and y are states such that*

$$I_0^{\varepsilon/2}(x) \geq I_\infty^{\varepsilon/2}(y),$$

then $x \xrightarrow{\varepsilon\text{-noisy}} y$.

Proof. We have already provided the proof for the $\varepsilon = 0$ case in Lemma 32. To extend the result to $\varepsilon > 0$, note that

$$I_0^{\varepsilon/2}(x) = \max_{\bar{x}: \mathcal{D}(\bar{x}, x) \leq \varepsilon/2} I_0(\bar{x}),$$

$$I_\infty^{\varepsilon/2}(y) = \min_{\bar{y}: \mathcal{D}(\bar{y}, y) \leq \varepsilon/2} I_\infty(\bar{y}).$$

Let \bar{x} and \bar{y} be states that achieve the maximum respectively minimum in these optimizations. Then

$$I_0(\bar{x}) = I_0^{\varepsilon/2}(x) \geq I_\infty^{\varepsilon/2}(y) = I_\infty(\bar{y}).$$

Due to Lemma 32, there is a noisy operation N such that $N(\bar{x}) = \bar{y}$. Now let $y' := N(x)$. Then

$$\begin{aligned} \mathcal{D}(y', y) &\leq \mathcal{D}(y', \bar{y}) + \mathcal{D}(\bar{y}, y) \\ &= \mathcal{D}(N(x), N(\bar{x})) + \mathcal{D}(\bar{y}, y) \\ &\leq \mathcal{D}(x, \bar{x}) + \varepsilon/2 \leq \varepsilon. \end{aligned}$$

This proves the claim. ■

The previous lemma gives a sufficient condition for approximate convertibility. Similarly, we can obtain a necessary condition:

Lemma 64 *If x and y are states such that*

$$x \xrightarrow{\varepsilon\text{-noisy}} y,$$

then $I_\infty^{\varepsilon+\delta}(y) \leq I_\infty^\delta(x)$ for every $\delta \geq 0$.

Proof. Denote the state which is ε -close to y and generated from x by y' . Lemma 59 implies that $I_\infty^\delta(x) \geq I_\infty^\delta(y')$. Let \bar{y} be any state such that $I_\infty^\delta(y') = I_\infty^\delta(\bar{y})$ and $\mathcal{D}(y', \bar{y}) \leq \delta$. Then $\mathcal{D}(y, \bar{y}) \leq \mathcal{D}(y, y') + \mathcal{D}(y', \bar{y}) \leq \varepsilon + \delta$, and we obtain $I_\infty^{\varepsilon+\delta}(y) \leq I_\infty(\bar{y}) = I_\infty^\delta(y') \leq I_\infty^\delta(x)$. ■

D. Asymptotic state conversion

Our previous results on single-shot ε -noisy nonuniformity distillation and dilution, Lemmas 63 and 64, allow us to recover (and slightly generalize) the result in (Horodecki *et al.*, 2003a,b) on the asymptotic conversion rate between two nonuniform states. Due to

Lemma 55, it is again sufficient to consider conversion between classical probability distributions.

The question we are interested in is the following. Suppose we are given n copies of some state x , which we would like to use to create as many copies of another state y as possible, using only noisy operations. That is, we would like to transform $x^{\otimes n}$ into $y^{\otimes m}$, with m as large as possible. However, we do not demand that the transformation is perfect—instead, we allow a small error ε . That is, for every n , we ask for the maximal integer m_n such that

$$x^{\otimes n} \xrightarrow{\varepsilon_n\text{-noisy}} y^{\otimes m_n},$$

and $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$. In particular, we are interested in how m_n scales with n . We will first answer this question for the case of constant error, i.e. $\varepsilon_n = \varepsilon$ for all n ; afterwards, we will obtain the answer for asymptotically vanishing error $\varepsilon_n \rightarrow 0$ as a simple corollary.

We have previously obtained necessary and sufficient conditions for approximate state conversion in terms of smooth order-0 and order- ∞ Rényi entropies. It is a well-known fact that these entropies converge to the Shannon entropy in the asymptotic limit of many copies of a state:

Lemma 65 *If the reference metric is taken to be either the purified distance \mathcal{D}_p from (6.2) or the trace distance \mathcal{D}_{tr} from (6.1), then we have for every $0 < \varepsilon < 1$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} H_\infty^\varepsilon(x^{\otimes n}) = H(x) \quad (6.6)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} H_0^\varepsilon(x^{\otimes n}) = H(x) \quad (6.7)$$

for all finite probability distributions x , where $H(x) = -\sum_i x_i \log x_i$ is the Shannon entropy.

Several versions of this statement can be found in the literature; for example, in (Tomamichel, 2012). Some care has to be taken, however, with respect to the slightly different definitions of the smooth entropies. For example, in contrast to (Tomamichel, 2012), we are here only considering classical probability distributions, we define smooth entropies in terms of optimization over *normalized* states (not subnormalized ones), and we define the smooth max-entropy as H_0^ε and not as $H_{1/2}^\varepsilon$, with H_α the Rényi entropy. For completeness, we give a proof of Lemma 65 in Appendix A.

In contrast with Section VI.C, where we allowed an arbitrary choice of contractive metric, the previous lemma is only proven for the trace distance and the purified distance. This is not completely unexpected: it simply cannot be true for all contractive metrics; for example, it fails for the discrete metric

$$\mathcal{D}(x, y) := \begin{cases} 0 & \text{if } x = y, \\ 1 & \text{if } x \neq y. \end{cases}$$

Furthermore, the trace distance and purified distance are the most frequently used metrics with clear operational

meaning. We leave it open whether the lemma above – or the following one – can be proven in greater generality.

Now we recover the result in (Horodecki *et al.*, 2003a,b) on the asymptotic conversion rate. Our proof turns out to be considerably simpler than the original one: it suffices to combine the single-shot results, Lemmas 63 and 64, with the asymptotic equipartition property, Lemma 65, to obtain the rate directly.

Lemma 66 *Let x and y be two states of possibly different dimensionalities that are not both uniform, let $0 < \varepsilon < 1$, and choose either the trace distance or the purified distance as reference metric. For every $n \in \mathbb{N}$, let m_n be the largest integer such that*

$$x^{\otimes n} \xrightarrow{\varepsilon\text{-noisy}} y^{\otimes m_n}.$$

Then

$$\lim_{n \rightarrow \infty} \frac{m_n}{n} = \frac{I(x)}{I(y)},$$

where $I(z) := \log d_z - H(z)$, with H the Shannon entropy. Similarly, let k_n be the smallest integer such that

$$x^{\otimes k_n} \xrightarrow{\varepsilon\text{-noisy}} y^{\otimes n}.$$

Then

$$\lim_{n \rightarrow \infty} \frac{k_n}{n} = \frac{I(y)}{I(x)}.$$

Remark. Due to Lemma 55, the analogous statement for quantum states follows immediately. Thus, we fully recover the result in (Horodecki *et al.*, 2003a,b), generalized to constant error ε .

Proof. A moment's thought shows that the statement of the lemma is trivially true if either x or y is uniform (such that $I(x) = 0$ or $I(y) = 0$); hence we only have to prove the case that $I(x) > 0$ and $I(y) > 0$. We only prove the first claim; the proof of the second claim is analogous. Clearly m_n is increasing in n . Since for every $m \in \mathbb{N}$, we have

$$I_0^{\varepsilon/2}(x^{\otimes n}) = n \cdot I(x) + o(n) > I_\infty^{\varepsilon/2}(y^{\otimes m})$$

for n large enough, Lemma 63 implies that $m_n \rightarrow \infty$ as $n \rightarrow \infty$. After some rearranging, Lemma 63 also says that if the inequality

$$\frac{m}{n} \leq \frac{\frac{1}{n} I_0^{\varepsilon/2}(x^{\otimes n})}{\frac{1}{m} I_\infty^{\varepsilon/2}(y^{\otimes m})}$$

is satisfied, then noisy operations can transform $x^{\otimes n}$ into a state that is ε -close to $y^{\otimes m}$. By construction, this is impossible for $m := m_n + 1$, so

$$\frac{m_n + 1}{n} > \frac{\frac{1}{n} I_0^{\varepsilon/2}(x^{\otimes n})}{\frac{1}{m_n + 1} I_\infty^{\varepsilon/2}(y^{\otimes (m_n + 1)})}.$$

Thus, we obtain

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{m_n}{n} &\geq \liminf_{n \rightarrow \infty} \frac{\frac{1}{n} I_0^{\varepsilon/2}(x^{\otimes n})}{\frac{1}{m_n+1} I_\infty^{\varepsilon/2}(y^{\otimes(m_n+1)})} \\ &= \frac{\log d_x - \lim_{n \rightarrow \infty} \frac{1}{n} H_0^{\varepsilon/2}(x^{\otimes n})}{\log d_y - \lim_{m \rightarrow \infty} \frac{1}{m} H_\infty^{\varepsilon/2}(y^{\otimes m})} \\ &= \frac{\log d_x - H(x)}{\log d_y - H(y)}. \end{aligned}$$

Conversely, choose $\delta \in (0, 1 - \varepsilon)$ arbitrarily (such that $\varepsilon + \delta < 1$), then Lemma 64 shows that the inequality

$$\frac{m_n}{n} \leq \frac{\frac{1}{n} I_\infty^\delta(x^{\otimes n})}{\frac{1}{m_n} I_\infty^{\varepsilon+\delta}(y^{\otimes m_n})}.$$

holds for every n , from which we can analogously infer by computing the limit of the right-hand side that

$$\limsup_{n \rightarrow \infty} \frac{m_n}{n} \leq \frac{\log d_x - H(x)}{\log d_y - H(y)}.$$

This proves that the limit $\lim_{n \rightarrow \infty} m_n/n$ exists and equals $I(x)/I(y)$. ■

As a simple corollary, we recover the result for *asymptotically vanishing error*, i.e. for the case that $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$ (however slowly). The proof idea is simple: use the method from Lemma 66 to achieve fixed error ε ; then, as n grows, decrease ε to zero, but do this slowly enough such that the fraction m_n/n still has the same limit as in Lemma 66. In the appendix, we give a proof that formally shows that this idea works.

By a *protocol*, we mean any choice of integers $(m_n)_{n \in \mathbb{N}}$ and noisy operations that transform $x^{\otimes n}$ into $y^{\otimes m_n}$ up to error ε_n for every n . With this definition, we can formalize the above idea into

Lemma 67 *There is a protocol for transforming n copies of x into m_n copies of y with asymptotically vanishing error at rate*

$$r := \lim_{n \rightarrow \infty} \frac{m_n}{n} = \frac{I(x)}{I(y)};$$

however, no higher rate is achievable by any protocol of this kind.

The proof is provided in Appendix A.

E. Nonuniformity cost and yield of asymptotic state conversion

Now we return to the problem posed in Subsection V.C, which is to quantify the nonuniformity cost and yield of state conversion. We can analyze this problem in the asymptotic case, and the answer turns out to be simple. Suppose that $I(x) < I(y)$, then – for n large enough – the approximate state conversion

$$x^{\otimes n} \xrightarrow{\varepsilon\text{-noisy}} y^{\otimes n} \quad (6.8)$$

is impossible – otherwise we would have $I(x) \geq I(y)$ according to Lemma 66 above. However, we may be able to create $y^{\otimes n}$ if, in addition to consuming $x^{\otimes n}$, we can consume a certain number, m_n , of pure bits (i.e. sharp states of nonuniformity 1), that is,

$$x^{\otimes n} \otimes s_1^{\otimes m_n} \xrightarrow{\varepsilon\text{-noisy}} y^{\otimes n}.$$

It is natural to try and determine the minimum value of m_n such that one can still achieve this conversion.

Similarly, if $I(x) > I(y)$, then the conversion (6.8) is possible, but “inefficient” – in addition to $y^{\otimes n}$, it will be possible to extract a certain number m_n of pure bits, such that

$$x^{\otimes n} \xrightarrow{\varepsilon\text{-noisy}} y^{\otimes n} \otimes s_1^{\otimes m_n}.$$

It is also natural to ask what the maximal value of m_n is such that this conversion can be achieved.

In Subsection V.C, we answered the analogous questions for the case of single states x and y (not $x^{\otimes n}$ or $y^{\otimes n}$) and for exact conversion. We provide the answer in the asymptotic case, for approximate conversion, in Lemma 68 below.

There are two natural ways to prove Lemma 68. One is to directly generalize the proof of Lemma 66, and to use the inequalities (for $\varepsilon, \delta \geq 0$)

$$\begin{aligned} I_\infty^{\varepsilon+\delta}(x \otimes y) &\leq I_\infty^\varepsilon(x) + I_\infty^\delta(y), \\ I_0^{\varepsilon+\delta}(x \otimes y) &\geq I_0^\varepsilon(x) + I_0^\delta(y) \end{aligned}$$

(following from analogous inequalities for smooth entropies) to split information measures of tensor products of states into parts. However, there is a simpler proof which establishes Lemma 68 as a corollary of Lemma 66 directly. It will be given below.

Lemma 68 (Asymptotic yield/cost of conversion)

Let x and y be two states of possibly different dimensionalities, let $0 < \varepsilon < 1$, and choose either the trace distance or the purified distance as reference metric. Consider the following two cases, characterized by the nonuniformity monotone $I(z) := \log d_z - H(z)$, with H the Shannon entropy:

- *Suppose that $I(x) \leq I(y)$. Let m_n be the smallest integer such that*

$$x^{\otimes n} \otimes s_1^{\otimes m_n} \xrightarrow{\varepsilon\text{-noisy}} y^{\otimes n}.$$

Then $\lim_{n \rightarrow \infty} \frac{m_n}{n} = I(y) - I(x)$.

- *Suppose that $I(x) \geq I(y)$. Let m_n be the largest integer such that*

$$x^{\otimes n} \xrightarrow{\varepsilon\text{-noisy}} y^{\otimes n} \otimes s_1^{\otimes m_n}.$$

Then $\lim_{n \rightarrow \infty} \frac{m_n}{n} = I(x) - I(y)$.

Proof. If x or y is uniform, the statement of the lemma follows directly from Lemma 66; thus, we will assume that $I(x) > 0$ and $I(y) > 0$. We give the proof of the first case, $I(x) \leq I(y)$, only; the proof of the second case is analogous. Denote by k_n the largest integer such that

$$x^{\otimes n} \xrightarrow{\varepsilon/2\text{-noisy}} s_1^{\otimes k_n}. \quad (6.9)$$

According to Lemma 66, we have $\lim_{n \rightarrow \infty} k_n/n = I(x)$. Now let l_n be the smallest integer such that

$$s_1^{\otimes l_n} \xrightarrow{\varepsilon/2\text{-noisy}} y^{\otimes n}. \quad (6.10)$$

According to Lemma 66, we have $\lim_{n \rightarrow \infty} l_n/n = I(y)$. Set $m := \max\{l_n - k_n, 0\}$. From (6.9), we obtain

$$x^{\otimes n} \otimes s_1^{\otimes m} \xrightarrow{\varepsilon/2\text{-noisy}} s_1^{\otimes (k_n+m)}.$$

Since $k_n + m \geq l_n$, we can first perform this conversion (and possibly discard some pure bits) and then perform conversion (6.10), in total yielding the conversion

$$x^{\otimes n} \otimes s_1^{\otimes m} \xrightarrow{\varepsilon\text{-noisy}} y^{\otimes n}.$$

Thus $m_n \leq m \leq |l_n - k_n|$. Dividing by n and taking the lim sup of both sides yields

$$\limsup_{n \rightarrow \infty} \frac{m_n}{n} \leq I(y) - I(x).$$

If $I(x) = I(y)$ then

$$\liminf_{n \rightarrow \infty} \frac{m_n}{n} \geq I(y) - I(x) \quad (6.11)$$

is trivially true. Thus, assume $I(x) < I(y)$, then it is clear that $m_n \rightarrow \infty$ as $n \rightarrow \infty$. Choose $\delta > 0$ such that $\varepsilon + \delta < 1$, and let i_n be the smallest integer such that

$$x^{\otimes i_n} \xrightarrow{\delta\text{-noisy}} s_1^{\otimes n}.$$

According to Lemma 66, we have $\lim_{n \rightarrow \infty} i_n/n = 1/I(x)$. Furthermore, let j_n be the smallest integer such that

$$x^{\otimes j_n} \xrightarrow{(\varepsilon + \delta)\text{-noisy}} y^{\otimes n}.$$

Due to Lemma 66, we have $\lim_{n \rightarrow \infty} j_n/n = I(y)/I(x)$. Moreover, we have the chain of conversions

$$\begin{aligned} x^{\otimes (n+i_{m_n})} &= x^{\otimes n} \otimes x^{\otimes i_{m_n}} \xrightarrow{\delta\text{-noisy}} x^{\otimes n} \otimes s_1^{\otimes m_n} \\ &\xrightarrow{\varepsilon\text{-noisy}} y^{\otimes n}, \end{aligned}$$

hence

$$x^{\otimes (n+i_{m_n})} \xrightarrow{(\varepsilon + \delta)\text{-noisy}} y^{\otimes n}.$$

By definition of j_n , it follows that $n + i_{m_n} \geq j_n$, and so

$$1 + \underbrace{\frac{i_{m_n}}{m_n}}_{\rightarrow 1/I(x)} \cdot \frac{m_n}{n} \geq \underbrace{\frac{j_n}{n}}_{\rightarrow I(y)/I(x)},$$

where the given limits are for $n \rightarrow \infty$. This establishes (6.11). ■

F. Approximate catalysis

Generalizing the notion of exact catalysis (discussed in Subsection V.D), one may ask whether a given state x can be converted into another state y *approximately* via the help of a catalyst in state z , where the notion of approximation is the following: the final state is ε -close to $y \otimes z$. It follows that the marginal on the system of this final state might not be precisely y , the marginal on the catalyst of this final state might not be precisely z , and the system and catalyst might in fact have some correlation in the final state. We ask under what conditions the “approximate noisy-trumping”

$$x \otimes z \xrightarrow{\varepsilon\text{-noisy}} y \otimes z \quad (6.12)$$

is possible. This problem has recently been studied in (Brandão *et al.*, 2015); here we give a brief summary of their results, translated into our notation. In this subsection, the reference metric will always taken to be the trace distance, $\mathcal{D} = \mathcal{D}_{\text{tr}}$, as defined in (6.1).

The first insight from (Brandão *et al.*, 2015) is related to the *embezzling* phenomenon (van Dam and Hayden, 2003). Suppose we fix $\varepsilon > 0$, and ask for what pairs of states x, y there exists some z such that the conversion (6.12) is possible. The answer turns out to be: *for all pairs of states*. That is, the approximate noisy-trumping relation in (6.12) is trivial (every state can be converted to every other) if ε is a fixed positive number which is independent of the states and catalyst.

The trick to show this is to use a huge-dimensional, unphysical catalyst z (an “embezzling state”) which has the property that one can distill an arbitrarily large amount of nonuniformity from it, while at the same time leaving it almost unmodified in trace distance. The intuitive reason why this is possible is as follows: *if two states z, z' are very close in the sense that $\mathcal{D}(z, z') \leq \varepsilon$, their nonuniformity content can still differ by arbitrarily large amounts*. This is easy to see, for example, in the case of Shannon nonuniformity I . If $d_z = d_{z'}$, we have by the Fannes inequality (Fannes, 1973; Nielsen and Chuang, 2010), as improved by Audenaert (Audenaert, 2007),

$$I(z') - I(z) = H(z) - H(z') \leq \varepsilon \log(d_z - 1) + h(\varepsilon)$$

if $\mathcal{D}(z', z) \leq \varepsilon$, where $h(\varepsilon) := -\varepsilon \log \varepsilon - (1 - \varepsilon) \log(1 - \varepsilon)$. The sharpness of this bound shows that for large d_z , states that are close can still have significantly different amounts of nonuniformity.

However, this inequality also gives a hint on how to obtain a more interesting notion of approximate noisy-trumping: *make sure that $\varepsilon \log(d_z - 1)$ is bounded, by decreasing ε with d_z* . This prescription, interpreted as quantifying the “error per particle” in (Brandão *et al.*, 2015), leads to a notion of approximate noisy-trumping that is fully characterized by the Shannon nonuniformity I :

Theorem 69 (Thm. 14 in (Brandão *et al.*, 2015))
Let x, y be distributions with $d_x = d_y =: d$. If there

exists a catalyst z such that

$$x \otimes z \xrightarrow{(\varepsilon/\log d_z)\text{-noisy}} y \otimes z,$$

then

$$I(x) \geq I(y) - 2\varepsilon - \frac{2\varepsilon \log d}{\log d_z} - h\left(\frac{2\varepsilon}{\log d_z}\right).$$

Conversely if $I(x) > I(y)$, then for all sufficiently large d_z there exists a catalyst z such that

$$x \otimes z \xrightarrow{\delta\text{-noisy}} y \otimes z, \quad \text{where } \delta = \exp\left(-\Omega(\sqrt{\log d_z})\right).$$

Roughly speaking, the theorem above states that x can be converted to y under approximate catalytic noisy operations if and only if the Shannon nonuniformity of x is bigger than that of y . On the other hand, we know from Lemma 46 that x can be converted to y under *exact* catalytic noisy operations if and only if for all $p \geq 0$ the Rényi p -nonuniformity of x is bigger than that of y . The reason there is no conflict between these two results is that all the Rényi p -nonuniformities with $p \neq 1$ are *not* asymptotically continuous.

VII. CONCLUSIONS

We have focussed in this article on problems of state conversion under noisy operations. In the single-copy regime, we have studied exact and approximate state conversion, catalytic and noncatalytic. In each case, we have discussed the necessary and sufficient conditions for the conversion to be possible. These results have interesting consequences for the status of the second law of thermodynamics.

The standard formulation of the second law is that entropy does not decrease. This is clearly inadequate as a criterion for the possibility of exact state conversion, with or without a catalyst, because evaluating the value of a single monotone can only generate a *total order* over states while noisy operations, catalytic or noncatalytic, induce a quasi-order.

Fortunately, one can rehabilitate the second law as a criterion for exact state conversion. To decide on the possibility of noncatalytic state conversion, one can compare the Lorenz curves of the states. If one wants to make the decision by finite means, one can use either one of the two state conversion witnesses described in Section V.E. To decide on the possibility of state conversion in the presence of a catalyst, one can compare the set of order- p Rényi nonuniformities for the states.

The question of whether the second law is adequate as a criterion for *approximate* state conversion is more subtle. Theorem 69 shows that under a particular notion of approximate catalysis, wherein the final state of the system and the catalyst must be $\varepsilon/\log d_z$ -close to the target (where d_z is the dimension of the catalyst), a state x can be converted to y (of equal dimension) if and only if the

entropy of y is greater than that of x . Leaving aside the case of unequal dimensions, this appears to be a vindication of the standard formulation of the second law. This notion of approximate catalytic state conversion, however, (like the one wherein the final state is required to be ε -close to the target state for some fixed ε) resembles the phenomenon of embezzlement insofar as a significant amount of nonuniformity is drawn from the catalyst.

A proper notion of approximate catalytic conversion ought to have the feature that the nonuniformity required to form y comes from x and not from the catalyst. The catalyst might still degrade with use, but one could require that the number of times one can reuse the catalyst is independent of the nature of the state conversions that it has facilitated. One can, for instance, define a notion of approximate catalytic conversion as follows: the final state must be such that by consuming some additional small amount of nonuniformity, one can convert it to the desired target state of the system together with the catalyst. Essentially, one is then using the state conversion witness Λ from Definition 35 to define a metric over the states. A similar definition can be made with the catalytic state conversion witness Λ_{cat} from Definition 48. This sort of notion was also proposed in (Brandão *et al.*, 2015). For any given finite amount of nonuniformity ε that one uses to define the degree of approximation, one can always find a pair of incomparable states x and y such that $|\Lambda(x||y)| \geq \varepsilon$ and $|\Lambda(y||x)| \geq \varepsilon$, and therefore under this notion of approximate catalytic conversion, the order over states remains a quasi-order. Necessarily then, the criterion for state conversion cannot be expressed in terms of the value of a single measure of nonuniformity, and the standard formulation of the second law is again inadequate.

Another set of results that we have described concerns the cost and yield of nonuniformity for state preparations and state conversions, both single-shot and asymptotic, as well as catalytic and noncatalytic. We have introduced a 1-parameter family of states, the sharp states, and shown that these can provide a “gold standard” of nonuniformity, in terms of which we can measure the costs and yields. Sharp states can be used to simulate erasure operations, so the nonuniformity yield of some process determines the amount of erasure, hence “informational work”, that can be extracted from a state or a state conversion process.

Lemma 66 implies that the rate at which pure bits must be consumed to generate x is equal to the rate at which pure bits can be distilled from x , in the asymptotic limit, namely $I(x)$ (because the nonuniformity of a pure bit is 1). Lemmas 30 and 31, on the other hand, show that in the single-shot case there is a gap: the nonuniformity of sharp state that is needed to form a single copy of x is $I_\infty(x)$, while one can only distill a sharp state of nonuniformity $I_0(x)$, and in general, $I_0(x) < I_\infty(x)$. Furthermore, Corollaries 50 and 51 demonstrate that the amount of nonuniformity one requires to form a state and the amount one can distill from a state do not change in

the presence of a catalyst.

Turning to the nonuniformity costs and yields of state conversion, Lemma 68 shows that in the asymptotic limit, the cost per copy of taking x to y and the yield per copy of taking y back to x are both equal to $I(y) - I(x)$, hence the process is reversible. In the single-shot case, however, there is a gap between cost and yield. Propositions 36 and 37 imply that the nonuniformity cost of the conversion of x to y is $\Lambda(x||y)$, while the nonuniformity yield of the conversion of y back to x is $-\Lambda(y||x)$ (which is positive whenever $\Lambda(x||y)$ is negative). In general $|\Lambda(x||y)| \neq |\Lambda(y||x)|$, so the conversion is not reversible.

To summarize, what emerges from this analysis is that attempting state preparations or state conversions one copy at a time is inefficient. It is only by processing asymptotically many copies at once that one can achieve perfect efficiency.

We now consider some open questions that remain in the resource theory of nonuniformity.

Although we have talked about simulating erasure operations using sharp states, there is a more general question concerning the precise resource requirements for simulating *any* given non-noisy operation, or relatedly, what non-noisy operations can be simulated from a given nonuniform state. An example of such a problem is to find the set of one-qubit quantum channels that can be implemented using an ancilla that has at most one bit of nonuniformity. This problem has been considered by several authors (Terhal *et al.*, 1999; Zalka and Rieffel, 2002) and a complete characterization of the class of accessible one-qubit channels is now known (Narang and Arvind, 2007). It remains to solve the problem beyond the case of one-qubit systems and one-qubit ancillas.

Another set of open questions concerns the situation wherein the restriction to noisy operations is combined with a locality restriction, such that an agent only has access to part of a composite system. In this case, quantum state conversion problems *do not* reduce to classical state conversion problems: entanglement between the accessible and inaccessible parts makes the problem inherently quantum. There is some work already in this direction (Faist *et al.*, 2012).

For almost every question about nonuniformity discussed in this article, there is a simple analogue within the resource theory of athermality. For one, it is the case that for quantum states that are block-diagonal across energy eigenspaces, every state conversion problem reduces to the corresponding problem for classical statistical states. Furthermore, state conversion problems under classical thermal operations are simple generalizations of those same problems under noisy classical operations. The equivalence classes of states under classical thermal operations are associated with *Gibbs-rescaled* histograms (Egloff *et al.*, 2012), which plot the ratios x_i/q_i in decreasing order, with q the thermal state for the associated system. They can alternatively be associated with the cumulative integral of these functions, which

are the athermal analogues of Lorenz curves. The curve of x being everywhere greater than or equal to that of y is the condition for $x \mapsto y$ under thermal operations, as shown in (Horodecki and Oppenheim, 2013). Functions over states that are nonincreasing under thermal operations, i.e. *athermality monotones*, can be defined from geometric features of these curves or from Schur-convex functions relative to the thermal state. As a consequence, for every cousin of the Shannon entropy in the zoo of entropies, there is an analogous cousin of the thermodynamic free energy.

The fact that the reduction of quantum state conversion problems to classical ones holds only for states that are block-diagonal in the energy eigenspaces is perhaps the most significant manner in which the theory of athermality must go beyond the theory of nonuniformity. Indeed, it remains an open problem to identify the thermal quasi-order over *all* quantum states. Note, however, that because a state that has coherence between energy eigenspaces is asymmetric relative to time translations, these questions are informed by results in the resource theory of asymmetry (Bartlett *et al.*, 2007; Gour and Spekkens, 2008; Marvian and Spekkens, 2013). Indeed, if one has access to a reference frame for time (a clock), then by applying energy-conserving unitaries to the composite of system and clock, it is possible to simulate unitaries on the system that can rotate between the energy eigenbasis and other bases. In this sense, a clock can act as a kind of catalyst for these state conversions (Brandão *et al.*, 2013).

We end with some general comments on the use of the framework of resource theories in the study of thermodynamics.

The standard tradition of enquiry in physics is *dynamicist*: The physicist’s job is to describe the natural dynamical behaviour of a system, without reference to human agents or their purposes. There is, however, a complementary *agent-centric* approach which focuses instead on characterizing limitations on an agent’s control of the behaviour of a system, or questions about what sorts of inferences an agent can make about the system (Wallace, 2013). Thermodynamics partakes in both traditions. One can see this easily by comparing various statements of the second law. One that is clearly in the dynamicist tradition is Clausius’s original statement: “Heat can never pass from a colder to a warmer body without some other change, connected therewith, occurring at the same time” (Clausius, 1867). On the other hand, the version of the Kelvin-Planck statement that is found in most textbooks is clearly agent-centric: “it is impossible to devise a cyclically operating device, the sole effect of which is to absorb energy in the form of heat from a single thermal reservoir and to deliver an equivalent amount of work” (Rao, 1997).

The resource-theoretic approach to thermodynamics that has been explored in this article is clearly agent-centric. The question of whether some state conversion which can be achieved by the free operations also arises

under *natural dynamics* (that is, in the absence of agents bringing them about) is not one that a resource theory seeks to answer. Nonetheless, the resource-theoretic approach might be adaptable to such questions. Specifically, if one wishes to characterize a certain subset of the free operations as more difficult to realize than others (on the grounds that they *do not* arise naturally), then one ought to identify the features that distinguish these operations and redefine the set of free operations to exclude them.

VIII. ACKNOWLEDGEMENTS

The authors would like to thank Fernando Brandão, Oscar Dahlsten, Nilanjana Datta, Lídia del Rio, Jonathan Oppenheim and Joe Renes for discussions. Research at Perimeter Institute is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MRI. GG and VN are supported by the Government of Canada through NSERC. This work has been supported by the COST network.

Appendix A: Proofs

In the main text, the proofs of some lemmas have been deferred to the appendix. We provide those proofs here.

Lemma 5 *Noisy quantum operations are a strict subset of the unital operations and, in the case of equal dimension of input and output space, a strict superset of the mixtures of unitaries.*

Proof. We first demonstrate the inclusions, then show that they are strict. Noisy operations are necessarily unital because if the input state is completely mixed, then after adjoining an ancilla in a completely mixed state, implementing a unitary, and taking a partial trace, one is necessarily left with a completely mixed state in the output. Mixtures of unitaries are necessarily noisy operations because one can implement them as follows. Suppose the ensemble of unitaries is $\{p_i, U_i\}$. One prepares an ancilla of arbitrarily large dimension in the completely mixed state, partitions its Hilbert space into subspaces, the relative dimensions of which are described by the distribution $\{p_i\}$. Next, one implements a controlled unitary with the ancilla as the control and the system as the target, where, if the ancilla is found in the subspace i , the unitary U_i is implemented on the system. Such a controlled unitary is, of course, itself a unitary on the composite of system and ancilla. Finally, one discards the ancilla.

Showing each inclusion's strictness is much harder. An example of a unital operation that is not a noisy operation has been provided in (Haagerup and Musat, 2011). Meanwhile, the fact that not every noisy operation is a mixture of unitaries has been shown by Shor in (Shor,

2010), making use of a result in (Mendl and Wolf, 2009). We refer the reader to (Shor, 2010) for details. ■

Lemma 6 *The set of noisy classical operations coincides with the set of uniform-preserving stochastic matrices and, in the case of equal dimension of input and output spaces (where the uniform-preserving stochastic matrices are the doubly-stochastic matrices), it coincides with the set of mixtures of permutations.*

Proof. It is straightforward to see that noisy classical operations are stochastic matrices. The proof that they are uniform-preserving proceeds in precise analogy to the proof that the noisy quantum operations are unital: if the input distribution is uniform, then by adjoining an ancilla in a uniform state and implementing a permutation, one creates a uniform state on the whole space, and every marginal is then also a uniform state.

The converse direction, that every uniform-preserving stochastic matrix can be realized as a noisy operation, is the nontrivial one. When the input and output vector spaces are of equal dimension, the stochastic uniform-preserving matrices are doubly stochastic, and a famous result due to Birkhoff (Birkhoff, 1946) establishes that every doubly stochastic matrix is achievable as a mixture of permutations. Therefore, it suffices to show that every mixture of permutations can be realized as a noisy classical operation. To implement permutation P_i with probability p_i , prepare an arbitrarily large ancilla in the uniform state, and partition its sample space into subsets of relative size p_i . Next, implement a controlled permutation with the ancilla as the control and the system as the target, where, if the ancilla is found in the subset i , the permutation P_i is implemented on the system. Such a controlled permutation is, of course, itself a permutation on the composite of system and ancilla. Finally, discard the ancilla.

Finally, we can show that even if the input and output spaces have different dimensions, every uniform-preserving stochastic matrix can be achieved as a noisy classical operation.

Let D be a uniform-preserving stochastic matrix from a d_{in} -dimensional probability space to a d_{out} -dimensional one. We now append an ancillary system to the input and one to the output such that the two composites are of equal dimension. That is, we define ancillary systems of dimensions d_1 and d_2 such that $d_{\text{in}}d_1 = d_{\text{out}}d_2$. Next, we define D_0 to be the $d_2 \times d_1$ matrix all of whose entries are $1/d_2$. This is clearly just the stochastic matrix that maps every state of dimension d_1 to the uniform state of dimension d_2 and is consequently uniform-preserving. It follows that the $(d_{\text{out}}d_2) \times (d_{\text{in}}d_1)$ matrix $D \otimes D_0$ is also a uniform-preserving stochastic matrix. Given that $d_{\text{out}}d_2 = d_{\text{in}}d_1$, it follows that $D \otimes D_0$ is doubly-stochastic and hence can be implemented by a mixture of permutations.

It follows that, if D is any uniform-preserving stochastic matrix, then it can be implemented by first adjoining a uniform state of dimension d_1 (which is a noisy oper-

ation), implementing a mixture of permutations (which, as shown earlier in this proof, is a noisy operation), and finally marginalizing over the ancillary subsystem of dimension d_2 (which is also a noisy operation). Given that every step of the implementation is a noisy operation, the overall operation is, as well. ■

Lemma 7 *There exists a noisy quantum operation that achieves the quantum state conversion $\rho \mapsto \sigma$ if and only if there is a noisy classical operation that achieves the classical state conversion $\lambda(\rho) \mapsto \lambda(\sigma)$.*

Proof. The possibility of the quantum state interconversion implies the existence of a quantum operation \mathcal{E} such that

$$\sigma = \mathcal{E}(\rho). \quad (\text{A1})$$

Denoting the j th eigenvector of σ by $|\phi_j\rangle$ and the k th eigenvector of ρ by $|\psi_k\rangle$, we have

$$\sum_k \lambda_k(\sigma) |\psi_k\rangle\langle\psi_k| = \mathcal{E} \left(\sum_j \lambda_j(\rho) |\phi_j\rangle\langle\phi_j| \right), \quad (\text{A2})$$

or equivalently,

$$\lambda_k(\sigma) = \sum_j \langle\psi_k| \mathcal{E}(|\phi_j\rangle\langle\phi_j|) |\psi_k\rangle \lambda_j(\rho). \quad (\text{A3})$$

If we define the matrix D by

$$D_{kj} = \langle\psi_k| \mathcal{E}(|\phi_j\rangle\langle\phi_j|) |\psi_k\rangle, \quad (\text{A4})$$

then we can write Eq. (A3) as

$$\lambda(\sigma) = D\lambda(\rho). \quad (\text{A5})$$

We seek to show that \mathcal{E} is a noisy quantum operation if and only if D is a noisy classical operation. The forward implication follows from the fact that

$$\begin{aligned} \sum_{j=1}^{d_{\text{in}}} D_{kj} &= \langle\psi_k| \mathcal{E}(I_{\text{in}}) |\psi_k\rangle = \langle\psi_k| \frac{d_{\text{in}}}{d_{\text{out}}} I_{\text{out}} |\psi_k\rangle = \frac{d_{\text{in}}}{d_{\text{out}}} \\ \sum_{k=1}^{d_{\text{out}}} D_{kj} &= \sum_k \text{Tr} [|\psi_k\rangle\langle\psi_k| \mathcal{E}(|\phi_j\rangle\langle\phi_j|)] \\ &= \text{Tr} (\mathcal{E}(|\phi_j\rangle\langle\phi_j|)) = \text{Tr} |\phi_j\rangle\langle\phi_j| = 1, \end{aligned}$$

where we have used the fact that \mathcal{E} is unital and the fact that \mathcal{E} is trace-preserving. This implies that D is a uniform-preserving stochastic matrix, which, by Lemma 6, implies that it is a noisy classical operation.

The reverse implication follows from the fact that we can define \mathcal{E} by

$$\mathcal{E}(\cdot) = \sum_{k,j} D_{kj} |\psi_k\rangle\langle\phi_j| (\cdot) |\phi_j\rangle\langle\psi_k|. \quad (\text{A6})$$

By assumption, D is a noisy classical operation, so it can be implemented by adjoining an ancilla in the uniform

state, performing a permutation on the composite, and summing over a subsystem. Let i be an index for the sample space of the ancilla, and let l be an index for the sample space of the subsystem that is summed over. Let $[R]_{kl,ji}$ denote the matrix elements of the permutation. It follows from Eq. (2.2) that $D_{kj} = (1/d_a) \sum_{l,i} [R]_{kl,ji}$, where we have used the fact that $[m_a]_i = 1/d_a$, with d_a the ancilla's dimension. Letting $\{|\mu_i\rangle\}$ denote an orthogonal set of vectors for the ancilla a , and $\{|\nu_l\rangle\}$ for the subsystem \bar{a} that is summed over, we have

$$\mathcal{E}(\rho) \quad (\text{A7})$$

$$\begin{aligned} &= \frac{1}{d_a} \sum_{k,j} \sum_{l,i} [R]_{kl,ji} |\psi_k\rangle\langle\phi_j| (\rho) |\phi_j\rangle\langle\psi_k| \\ &= \text{Tr}_{\bar{a}} \left[\sum_{klji} [R]_{kl,ji} |\psi_k\rangle\langle\psi_k| |\nu_l\rangle\langle\nu_l| \langle\phi_j| (\rho \otimes \frac{I_a}{d_a}) |\phi_j\rangle\langle\phi_j| |\mu_i\rangle\langle\mu_i| |\nu_l\rangle\langle\nu_l| \right] \end{aligned} \quad (\text{A8})$$

$$= \text{Tr}_{\bar{a}} \left[V(\rho \otimes \frac{1}{d_a} I_a) V^\dagger \right], \quad (\text{A10})$$

where we have implicitly defined a unitary V . In this form, it is clear that \mathcal{E} is a noisy quantum operation. ■

Lemma 10 *Let $\rho \in \mathcal{L}(\mathcal{H}_{\text{in}})$ and $\sigma \in \mathcal{L}(\mathcal{H}_{\text{out}})$. Then, the following propositions are equivalent:*

- (i) $\rho \mapsto \sigma$ by a noisy operation
- (ii) $\rho \mapsto \sigma$ by a unital operation.

If ρ and σ are of equal dimension, then (i) and (ii) are also equivalent to

- (iii) $\rho \mapsto \sigma$ by a mixture of unitaries.

Proof. Clearly, (i) implies (ii) because a noisy operation is unital, and (iii) implies (i) because a mixture of unitaries is a noisy operation. The nontrivial implications are (ii) to (i) (for unequal dimension) and (ii) to (iii) (for equal dimensions). The proof that (ii) implies (i) can be inferred from the proof of Lemma 7, as follows. In the forward direction of that proof, we inferred from the existence of a noisy quantum operation \mathcal{E} taking ρ to σ that there was a noisy classical operation taking the spectrum of ρ to the spectrum of σ . But that inference only relied on the fact that \mathcal{E} was unital. Then, in the reverse direction of that proof, we showed that from a noisy classical operation taking the spectrum of ρ to the spectrum of σ , one can construct a noisy quantum operation taking ρ to σ . Therefore, (ii) implies (i). In the case of equal dimension, we simply note that the noisy classical operation can be written as a mixture of permutations (by Birkhoff's Theorem (Birkhoff, 1946)), and hence the noisy quantum operation that one constructs from it is a mixture of unitaries. Therefore, (ii) implies (iii). ■

Lemma 58 *Let \mathcal{D} be any contractive metric on the quantum states for which there exists a norm $\|\cdot\|$ on the self-adjoint matrices such that $\mathcal{D}(\rho, \sigma) = \|\rho - \sigma\|$ (for example*

the trace distance, $\mathcal{D} = \mathcal{D}_{\text{tr}}$ with $\|\cdot\| = \frac{1}{2}\|\cdot\|_1$). Then the quantum version of the smooth max-entropy,

$$S_0^\varepsilon(\rho) := \min_{\bar{\rho}: \mathcal{D}(\rho, \bar{\rho}) \leq \varepsilon} S_0(\bar{\rho}),$$

with $S_0(\bar{\rho}) := \log \text{rank}(\bar{\rho})$, agrees with the classical smooth max-entropy on the spectrum of ρ :

$$S_0^\varepsilon(\rho) = H_0^\varepsilon(\lambda(\rho)).$$

Proof. In the following, we will choose a basis such that ρ is diagonal; in particular, $\rho = \text{diag}(r_1, r_2, \dots, r_n)$ with $r_1 \geq r_2 \geq \dots \geq r_n$. Let s be any distribution such that $\mathcal{D}(\lambda(\rho), s) \leq \varepsilon$ and $H_0^\varepsilon(\lambda(\rho)) = H_0(s)$. Let $\sigma := \text{diag}(s_1, s_2, \dots, s_n)$, then $\mathcal{D}(\rho, \sigma) = \mathcal{D}(\lambda(\rho), s) \leq \varepsilon$. Thus

$$S_0^\varepsilon(\rho) \leq S_0(\sigma) = H_0(s) = H_0^\varepsilon(\lambda(\rho)).$$

To prove the converse inequality, let σ be any quantum state with $\mathcal{D}(\rho, \sigma) \leq \varepsilon$ and $S_0^\varepsilon(\rho) = S_0(\sigma)$. Denote by $s = (s_1, s_2, \dots, s_n)$ the eigenvalues of σ in non-increasing order, i.e. $s_1 \geq s_2 \geq \dots \geq s_n$. Contractivity of the metric implies unitary invariance of the norm. Thus, it follows from (Bhatia, 1997, Ineq. (IV.62)) that $\mathcal{D}(\lambda(\rho), s) \leq \mathcal{D}(\rho, \sigma) \leq \varepsilon$. Hence

$$H_0^\varepsilon(\lambda(\rho)) \leq H_0(s) = S_0(\sigma) = S_0^\varepsilon(\rho).$$

This proves the claim. ■

Unfortunately, the purified distance does not satisfy the premise of Lemma 58: it does not come from a norm.

Lemma 60 *If the trace distance \mathcal{D}_{tr} is chosen as reference metric, then any distribution x has smooth Rényi 0-entropy*

$$H_0^\varepsilon(x) = \log k,$$

where $k \geq 1$ is chosen as the smallest integer such that $\sum_{i=1}^k x_i^\downarrow \geq 1 - \varepsilon$, with $x_1^\downarrow \geq x_2^\downarrow \geq \dots$ denoting the components of x in non-increasing order.

Proof. Since $H_0^\varepsilon(x)$ is invariant with respect to permutations of the components of x , we may assume that the components of x are ordered, such that $x_i = x_i^\downarrow$. For a given state $x = (x_1, \dots, x_n)$, define k as described above, and set $\mathcal{N} := \sum_{i=1}^k x_i \geq 1 - \varepsilon$. Set $\tilde{x} := (x_1, \dots, x_k, 0, \dots, 0) \in \mathbb{R}^n$, then $x' := \tilde{x}/\mathcal{N}$ is a state. A simple calculation shows that $\mathcal{D}_{\text{tr}}(x, x') = 1 - \mathcal{N} \leq \varepsilon$, thus $H_0^\varepsilon(x) \leq H_0(x') = \log k$.

Conversely, let $y \in \mathbb{R}^n$ be any state with $H_0(y) < \log k$, then consider the support of y , $\text{supp}(y) := \{i \in \{1, \dots, n\} \mid y_i \neq 0\}$. This has cardinality $|\text{supp}(y)| < k$, thus $\sum_{i \in \text{supp}(y)} x_i < 1 - \varepsilon$ by the construction of k . Recalling an alternative definition of the classical trace distance [(Nielsen and Chuang, 2010), eq. (9.4)] and comparing the values of x and y on the kernel of y , $\ker(y)$,

we get

$$\begin{aligned} \mathcal{D}_{\text{tr}}(x, y) &= \sup_S \left| \sum_{i \in S} x_i - \sum_{i \in S} y_i \right| \\ &\geq \left| \sum_{i \in \ker(y)} x_i - \sum_{i \in \ker(y)} y_i \right| \\ &= 1 - \sum_{i \in \text{supp}(y)} x_i > \varepsilon. \end{aligned}$$

This shows that $H_0^\varepsilon(x) \geq \log k$, proving the claim. ■

In (Tomamichel, 2012), it is shown that the purified distance upper-bounds the trace distance:

$$\mathcal{D}_{\text{tr}}(x, y) \leq \mathcal{D}_p(x, y). \quad (\text{A11})$$

This is the final ingredient to prove Lemma 65 from the main text, the asymptotic equipartition property:

Lemma 65 *If the reference metric is taken to be either the purified distance \mathcal{D}_p or the trace distance \mathcal{D}_{tr} , then we have for every $0 < \varepsilon < 1$*

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} H_\infty^\varepsilon(x^{\otimes n}) &= H(x) \\ \lim_{n \rightarrow \infty} \frac{1}{n} H_0^\varepsilon(x^{\otimes n}) &= H(x) \end{aligned}$$

for all finite probability distributions x , where $H(x) = -\sum_i x_i \log x_i$ is the Shannon entropy.

Proof. For the sake of this proof, let H_∞^ε and H_0^ε be the smooth entropies defined with respect to the trace distance \mathcal{D}_{tr} , and denote by $\bar{H}_\infty^\varepsilon$ and \bar{H}_0^ε the smooth entropies defined with respect to the purified distance \mathcal{D}_p . Denoting the ε -ball around some state x according to a metric \mathcal{D} by

$$B_{\mathcal{D}}^\varepsilon(x) := \{y \mid \mathcal{D}(x, y) \leq \varepsilon\},$$

inequality (A11) implies that

$$B_{\mathcal{D}_p}^\varepsilon(x) \subseteq B_{\mathcal{D}_{\text{tr}}}^\varepsilon(x). \quad (\text{A12})$$

We show our lemma directly via the asymptotic equipartition property as given, for example, in Theorem 3.1.2 of (Cover and Thomas, 2012). Given any distribution x , define the ε -typical set $A_\varepsilon^{(n)}$ of all sequences $s = (s_1, \dots, s_n)$ of length n as

$$A_\varepsilon^{(n)} := \{s \mid 2^{-n(H(x)+\varepsilon)} \leq x^{\otimes n}(s) \leq 2^{-n(H(x)-\varepsilon)}\}.$$

Then we can find a sequence $(\varepsilon_n)_{n \in \mathbb{N}}$ with $\varepsilon_n \xrightarrow{n \rightarrow \infty} 0$ such that the sequence of sets $A^{(n)} := A_{\varepsilon_n}^{(n)}$ satisfies

$$\begin{aligned} x^{\otimes n}(A^{(n)}) &> 1 - \varepsilon_n, \\ (1 - \varepsilon_n) 2^{n(H(x)-\varepsilon_n)} &\leq |A^{(n)}| \leq 2^{n(H(x)+\varepsilon_n)}. \end{aligned}$$

Let $\varepsilon \in (0, 1)$ be arbitrary, and fix any $\delta > 0$. We will now prove the following claim: if n is large enough, then any

distribution $q^{(n)}$ on n symbols with support on some set $Q^{(n)}$, where $|Q^{(n)}| \leq 2^{n(H(x)-\delta)}$, has $\mathcal{D}_{\text{tr}}(q^{(n)}, x^{\otimes n}) > \varepsilon$. If we have a distribution $q^{(n)}$ with these properties, then

$$q^{(n)}(A^{(n)} \setminus Q^{(n)}) = 0,$$

but

$$\begin{aligned} x^{\otimes n}(A^{(n)} \setminus Q^{(n)}) &= x^{\otimes n}(A^{(n)}) - x^{\otimes n}(A^{(n)} \cap Q^{(n)}) \\ &> 1 - \varepsilon_n - 2^{-n(H(x)-\varepsilon_n)} \cdot |Q^{(n)}| \\ &\geq 1 - \varepsilon_n - 2^{-n(\delta-\varepsilon_n)} \\ &\xrightarrow{n \rightarrow \infty} 1. \end{aligned}$$

Since the trace distance gives the maximal possible probability with which two distributions can be distinguished, this implies that $\mathcal{D}_{\text{tr}}(q^{(n)}, x^{\otimes n}) > \varepsilon$ for all distributions $q^{(n)}$ of this kind if n is large enough. Thus, denoting by $\text{supp } q^{(n)}$ the set of s with $q^{(n)}(s) > 0$, we obtain

$$\mathcal{D}_{\text{tr}}(x^{\otimes n}, q^{(n)}) \leq \varepsilon \Rightarrow |\text{supp } q^{(n)}| > 2^{n(H(x)-\delta)}.$$

if n is large enough, which also implies that $H_0(q^{(n)}) > n(H(x) - \delta)$, and so

$$H_0^\varepsilon(x^{\otimes n}) > n(H(x) - \delta).$$

Consequently,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} H_0^\varepsilon(x^{\otimes n}) \geq H(x) - \delta.$$

Since this is true for every $\delta > 0$, we obtain

$$\liminf_{n \rightarrow \infty} \frac{1}{n} H_0^\varepsilon(x^{\otimes n}) \geq H(x). \quad (\text{A13})$$

Furthermore, the inequality

$$\begin{aligned} \bar{H}_0^\varepsilon(y) &= \min_{\bar{y} \in B_{\mathcal{D}_p}^\varepsilon(y)} H_0(\bar{y}) \\ &\geq \min_{\bar{y} \in B_{\mathcal{D}_{\text{tr}}}^\varepsilon(y)} H_0(\bar{y}) = H_0^\varepsilon(y) \end{aligned}$$

implies that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \bar{H}_0^\varepsilon(x^{\otimes n}) \geq H(x). \quad (\text{A14})$$

Analogously, suppose that $r^{(n)}$ is a distribution that satisfies one of the two equivalent conditions

$$\max_s r^{(n)}(s) \leq 2^{-n(H(x)+\delta)} \Leftrightarrow H_\infty(r^{(n)}) \geq n(H(x) + \delta),$$

where the maximum is over all sequences s of length n . Then it follows

$$r^{(n)}(A^{(n)}) \leq 2^{-n(H(x)+\delta)} \cdot |A^{(n)}| \leq 2^{-n(\delta-\varepsilon_n)}$$

which tends to zero for $n \rightarrow \infty$, while $x^{\otimes n}(A^{(n)})$ tends to one. This shows that $\mathcal{D}_{\text{tr}}(x^{\otimes n}, r^{(n)}) > \varepsilon$ for all distributions $r^{(n)}$ of this kind if n is large enough. Thus,

$$\mathcal{D}_{\text{tr}}(x^{\otimes n}, r^{(n)}) \leq \varepsilon \Rightarrow H_\infty(r^{(n)}) < n(H(x) + \delta)$$

if n is large enough, and so

$$H_\infty^\varepsilon(x^{\otimes n}) < n(H(x) + \delta).$$

Consequently,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} H_\infty^\varepsilon(x^{\otimes n}) \leq H(x) + \delta.$$

Since this is true for every $\delta > 0$, we obtain

$$\limsup_{n \rightarrow \infty} \frac{1}{n} H_\infty^\varepsilon(x^{\otimes n}) \leq H(x). \quad (\text{A15})$$

Furthermore, the inequality

$$\begin{aligned} \bar{H}_\infty^\varepsilon(y) &= \max_{\bar{y} \in B_{\mathcal{D}_p}^\varepsilon(y)} H_\infty(\bar{y}) \\ &\leq \max_{\bar{y} \in B_{\mathcal{D}_{\text{tr}}}^\varepsilon(y)} H_\infty(\bar{y}) = H_\infty^\varepsilon(y) \end{aligned}$$

implies that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \bar{H}_\infty^\varepsilon(x^{\otimes n}) \leq H(x). \quad (\text{A16})$$

Now define a distribution $q^{(n)}$ by

$$q^{(n)}(s) := \begin{cases} x^{\otimes n}(s)/x^{\otimes n}(A^{(n)}) & \text{if } s \in A^{(n)} \\ 0 & \text{otherwise.} \end{cases}$$

A simple calculation shows that

$$\mathcal{D}_{\text{tr}}(q^{(n)}, x^{\otimes n}) = \frac{1}{2} \left(1 - x^{\otimes n}(A^{(n)})\right) \xrightarrow{n \rightarrow \infty} 0,$$

thus

$$H_0^\varepsilon(x^{\otimes n}) \leq H_0(q^{(n)}) = \log |A^{(n)}| \leq n(H(x) + \varepsilon_n)$$

if n is large enough. Thus

$$\limsup_{n \rightarrow \infty} \frac{1}{n} H_0^\varepsilon(x^{\otimes n}) \leq H(x). \quad (\text{A17})$$

Similarly,

$$\begin{aligned} H_\infty^\varepsilon(x^{\otimes n}) &\geq H_\infty(q^{(n)}) \\ &= \log x^{\otimes n}(A^{(n)}) - \log \max_{s \in A^{(n)}} x^{\otimes n}(s) \\ &> \log(1 - \varepsilon_n) - \log 2^{-n(H(x)-\varepsilon_n)}. \end{aligned}$$

This shows that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} H_\infty^\varepsilon(x^{\otimes n}) \geq H(x). \quad (\text{A18})$$

Furthermore, an elementary calculation shows that

$$\mathcal{D}_p(q^{(n)}, x^{\otimes n}) = \sqrt{1 - x^{\otimes n}(A^{(n)})} \xrightarrow{n \rightarrow \infty} 0,$$

and repeating the calculations above for the purified distance yields the inequalities

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \bar{H}_0^\varepsilon(x^{\otimes n}) \leq H(x), \quad (\text{A19})$$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \bar{H}_\infty^\varepsilon(x^{\otimes n}) \geq H(x). \quad (\text{A20})$$

The inequalities (A13), (A14), (A15), (A16), (A17), (A18), (A19), and (A20) prove the lemma. ■

Lemma 67 *There is a protocol for transforming n copies of x into m_n copies of y with asymptotically vanishing error at rate*

$$r := \lim_{n \rightarrow \infty} \frac{m_n}{n} = \frac{I(x)}{I(y)};$$

however, no higher rate is achievable by any protocol of this kind.

Proof. Since every protocol with asymptotically vanishing error is a special case of a protocol with constant error, Lemma 66 implies the second part of the statement, i.e. that no higher rate than $r := I(x)/I(y)$ is achievable.

Now we show how to construct a protocol with asymptotically vanishing error from a protocol with constant error. For every constant error $\varepsilon > 0$, denote the corresponding integer m_n from Lemma 66 by $m_n(\varepsilon)$. For every fixed $\varepsilon > 0$, we have

$$\lim_{n \rightarrow \infty} \frac{m_n(\varepsilon)}{n} = r.$$

Thus, for every integer $k \in \mathbb{N}$, there exists an integer \tilde{n}_k such that

$$\left| \frac{m_n(1/k)}{n} - r \right| < \frac{1}{k} \quad (\text{A21})$$

for all $n \geq \tilde{n}_k$. Let $n_k := k + \max_{1 \leq \ell \leq k} \tilde{n}_\ell$, then n_k is strictly increasing in k , and (A21) holds in particular for all $n \geq n_k$.

The protocol now is as follows. Given any integer n , find the k such that $n \in [n_k, n_{k+1} - 1]$, which is possible if n is large enough, i.e. $n \geq n_{k=1}$. Then convert $x^{\otimes n}$ into a state y' which is $(1/k)$ -close to $y^{\otimes m_n(1/k)}$, which is possible by definition of the sequence $m_n(\varepsilon)$. That is, we convert n copies of x approximately into m_n copies of y , where $m_n = m_n(1/k)$. Since

$$\frac{m_n}{n} = \frac{m_n(1/k)}{n} \in \left(r - \frac{1}{k}, r + \frac{1}{k} \right)$$

and $k \rightarrow \infty$ as $n \rightarrow \infty$, we have $\lim_{n \rightarrow \infty} m_n/n = r$. Furthermore, the error $\varepsilon_n = 1/k$ tends to zero as $n \rightarrow \infty$. ■

Appendix B: Strong noisy-trumping and $p < 0$ Rényi nonuniformities

In this section, as announced in Subsection V.D, we study a notion of noisy-trumping which is an alternative

to Definition 45. We do this here in the appendix because this alternative definition turns out to have undesirable mathematical properties, rendering its physical interpretation problematic. However, we still think that the discussion may be interesting within the broader question of how to define a proper notion of trumping, and also because it has been mentioned in (Brandão *et al.*, 2015).

In the discussion of Subsection V.D, the first idea for defining a notion of noisy-trumping that removed the discontinuity mentioned there was to relax the demand of producing the target state perfectly, and instead allow the target state to be obtained to arbitrary accuracy. The following definition formalizes this idea. In contrast to Definition 45, it will require the input state to be prepared perfectly.

Definition 70 (Strong noisy-trumping) *We say that a state x strongly noisy-trumps another state y if and only if there is a sequence of state $(y_n)_{n \in \mathbb{N}}$, all of the same dimension as y , which converges to y , i.e. $\lim_{n \rightarrow \infty} y_n = y$, and a sequence of catalysts $(z_n)_{n \in \mathbb{N}}$ such that*

$$x \otimes z_n \xrightarrow{\text{noisy}} y_n \otimes z_n \text{ for every } n \in \mathbb{N}.$$

Note that the sequence of catalysts does not have to converge; moreover, the catalysts z_n may have different dimensionalities (d_{z_n} may even grow unboundedly).

This is indeed a stronger notion of trumping: if x strongly noisy-trumps y , then x also noisy-trumps y in the sense of Definition 45. This is because if x strongly noisy-trumps y , then (according to the definition above) x noisy-trumps every y_n , and thus due to Corollary 47 x also noisy-trumps y . Below, we will see that noisy-trumping and strong noisy-trumping are strictly different notions of trumping; this will follow from the fact that the corresponding quasi-orders are characterized by different complete sets of monotones.

We start by giving an analog of Lemma 46 for strong noisy-trumping. The proof again relies on the results by Klimesh (2007) and Turgut (2007).

Lemma 71 (Conditions for strong noisy-trumping)

Let x and y be any pair of states. Then the strong noisy-trumping relation is characterized by two cases:

(i) *The state x does not contain zeros. Then x strongly noisy-trumps y if and only if*

$$I_p(x) \geq I_p(y) \text{ for all } p \in \mathbb{R}.$$

(ii) *The state x contains zeros. Then x strongly noisy-trumps y if and only if*

$$I_p(x) \geq I_p(y) \text{ for all } p \geq 0.$$

Proof. Suppose that x strongly noisy-trumps y ; that is, there is a sequence $(y_n)_{n \in \mathbb{N}}$ with $\lim_{n \rightarrow \infty} y_n = y$ and a

sequence of catalysts $(z_n)_{n \in \mathbb{N}}$ such that $x \otimes z_n \xrightarrow{\text{noisy}} y_n \otimes z_n$. Since every I_p is a nonuniformity monotone, we get

$$I_p(x) + I_p(z_n) = I_p(x \otimes z_n) \geq I_p(y_n \otimes z_n) = I_p(y_n) + I_p(z_n).$$

If $I_p(z_n)$ is finite, then we can cancel it from both sides of the inequality and we are done. The only possibility for $I_p(z_n)$ to *not* be finite is for $p < 0$ and for z_n to have components that are zero. However, by Corollary 43, if z_n catalyzes the conversion of x to y_n then so does $z_n^{(I_0(z_n))}$ and the latter has full support. One can therefore repeat the argument for $z_n^{(I_0(z_n))}$.

Thus $I_p(x) \geq I_p(y_n)$. If $p \neq 0$, then I_p is continuous, so $y_n \rightarrow y$ for $n \rightarrow \infty$ implies $I_p(y_n) \rightarrow I_p(y)$ (including the possibility that $I_p(y) = \infty$). Thus, we can take the limit of the inequality and conclude

$$I_p(x) \geq I_p(y) \text{ for all } p \in \mathbb{R} \setminus \{0\}.$$

Finally, we can take the limit $p \searrow 0$ of the latter inequality, which proves that it also holds for $p = 0$.

Now suppose that x does not contain zeros, and that $I_p(x) \geq I_p(y)$ for all $p \in \mathbb{R}$. We may assume that $d_x = d_y$ for the same reasons as were given in the proof of Lemma 46. We may also assume that y is not maximally mixed. We have to show that x strongly noisy-trumps y ; we will do so by using the results of Lemma 44. Translating our inequality for I_p into the notation of this lemma, we obtain that $f_p(x) \geq f_p(y)$ for all $p \neq 0$. If y contained zeros, then we would have $I_p(y) = \infty$ and $I_p(x) < \infty$ for $p < 0$ which would contradict $I_p(x) \geq I_p(y)$. Thus y has no zeros either. Thus, we can use the limit

$$\lim_{p \searrow 0} \frac{1}{p} I_p(z) = -\log d_z - \frac{1}{d_z} \sum_{i=1}^{d_z} \log z_i \equiv -\log d_z + \frac{1}{d_z} f_0(z)$$

to conclude that $f_0(x) \geq f_0(y)$ (again, the same trick as was used in (Brandão *et al.*, 2015)).

For $n \in \mathbb{N}$, set $y_n := (1 - 1/n)y + (1/n)m^{(d_y)}$, then y_n does not contain zeros. Since all f_p are strictly Schur-concave, we have $f_p(y_n) < f_p(y)$, and so

$$f_p(x) > f_p(y_n) \text{ for all } p \in \mathbb{R}.$$

Lemma 44 implies that there exists some catalyst z_n such that $x \otimes z_n \xrightarrow{\text{noisy}} y_n \otimes z_n$. Since this works for every $n \in \mathbb{N}$ and $\lim_{n \rightarrow \infty} y_n = y$, this shows that x strongly noisy-trumps y .

Finally, suppose that x contains zeros, and define y_n as above. If $p \leq 0$ then $\infty = f_p(x) > f_p(y_n)$ is automatically satisfied, and we only need to prove that $f_p(x) > f_p(y_n)$ for $p > 0$, which follows from $I_p(x) \geq I_p(y)$ for $p > 0$. ■

This result motivates the following definition.

Definition 72 For an arbitrary pair of states x and y , define

$$\Lambda_{\text{cat}}^{\text{strong}}(x||y) := \inf_{p \in S_x} (I_p(x) - I_p(y)),$$

where

$$S_x := \begin{cases} \mathbb{R}^+ & \text{if } x \text{ contains zeros,} \\ \mathbb{R} & \text{otherwise.} \end{cases}$$

If x contains zeros, then $\Lambda_{\text{cat}}^{\text{strong}}(x||y) = \Lambda_{\text{cat}}(x||y)$; in general, we have the inequality

$$\Lambda_{\text{cat}}^{\text{strong}}(x||y) \leq \Lambda_{\text{cat}}(x||y).$$

The following proposition is a reformulation of Lemma 71 in terms of this new quantity:

Proposition 73 For an arbitrary pair of states x and y , x strongly noisy-trumps y if and only if

$$\Lambda_{\text{cat}}^{\text{strong}}(x||y) \geq 0.$$

Equivalently, the function $\Lambda_{\text{cat}}^{\text{strong}}(x||y)$ is a complete witness for strong noisy-trumping.

The nonuniformity of formation and the distillable nonuniformity (as determined in Proposition 30 and 31) are not changed by catalysis in the sense of strong noisy-trumping. In other words, Corollaries 50 and 51 also hold for strong noisy-trumping and the proofs are exactly analogous. The move from noisy-trumping to strong noisy-trumping *does* make a difference, however, for the nonuniformity cost and yield of state conversion.

Proposition 74 If $\Lambda_{\text{cat}}^{\text{strong}}(x||y) > 0$, then x strongly noisy-trumps $y \otimes s_I$ if and only if

$$I \leq \Lambda_{\text{cat}}^{\text{strong}}(x||y).$$

Proof. The case $I = 0$ is just Proposition 73; thus, we may assume $I > 0$. First, suppose that x does not contain any zeros. But then $0 = I_0(x) < I_0(y \otimes s_I)$, and so x cannot strongly noisy-trump $y \otimes s_I$. In fact,

$$\begin{aligned} \Lambda_{\text{cat}}^{\text{strong}}(x||y) &= \inf_{p \in \mathbb{R}} (I_p(x) - I_p(y)) \leq I_0(x) - I_0(y) \\ &= -I_0(y) \leq 0 < I. \end{aligned}$$

Now consider the case that x contains zeros. Then

$$\begin{aligned} \Lambda_{\text{cat}}^{\text{strong}}(x||y \otimes s_I) &= \inf_{p \geq 0} (I_p(x) - I_p(y \otimes s_I)) \\ &= \inf_{p \geq 0} (I_p(x) - I_p(y)) - I \\ &= \Lambda_{\text{cat}}^{\text{strong}}(x||y) - I. \end{aligned}$$

According to Proposition 73, x trumps $y \otimes s_I$ if and only if this expression is non-negative, that is, if and only if $\Lambda_{\text{cat}}^{\text{strong}}(x||y) \geq I$. ■

Surprisingly, the nonuniformity cost of “strong” catalytic state conversion is not determined by $\Lambda_{\text{cat}}^{\text{strong}}$, but by the quantity Λ_{cat} from Definition 48:

Proposition 75 *If $\Lambda_{\text{cat}}(x||y) < 0$, then $x \otimes s_I$ strongly noisy-trumps y if and only if*

$$I \geq -\Lambda_{\text{cat}}(x||y).$$

Proof. Since $x \otimes s_I$ contains zeros, we have

$$\begin{aligned} \Lambda_{\text{cat}}^{\text{strong}}(x \otimes s_I||y) &= \inf_{p \geq 0} (I_p(x \otimes s_I) - I_p(y)) \\ &= I + \inf_{p \geq 0} (I_p(x) - I_p(y)) \\ &= I + \Lambda_{\text{cat}}(x||y). \end{aligned}$$

According to Proposition 73, $x \otimes s_I$ strongly noisy-trumps y if and only if this expression is non-negative, that is, if and only if $I + \Lambda_{\text{cat}}(x||y) \geq 0$. ■

In the case where both x and y do not contain zeros, such that $\Lambda_{\text{cat}}^{\text{strong}}(x||y) < \Lambda_{\text{cat}}(x||y)$ is possible, Propositions 73 and 75 show an undesirable discontinuity in the nonuniformity cost of catalytic state conversion in the case of strong noisy-trumping: $x \otimes s_I$ strongly noisy-trumps y if and only if

$$\begin{cases} I \geq -\Lambda_{\text{cat}}^{\text{strong}}(x||y) & \text{if } I = 0 \\ I \geq -\Lambda_{\text{cat}}(x||y) & \text{if } I > 0. \end{cases}$$

A discontinuity like this was not present in the definition of noisy trumping used in the main text (Definition 45).

As a final observation, Proposition 75 implies that two different possible definitions of noisy trumping are equivalent:

Lemma 76 *Let x and y be any pair of states. Then*

$$x \otimes s_\delta \text{ strongly noisy-trumps } y \text{ for all } \delta > 0 \quad (\text{B1})$$

if and only if x noisy-trumps y .

Proof. First, consider the case that $\Lambda_{\text{cat}}(x||y) < 0$. According to Proposition 49, this means that x does not noisy-trump y . Indeed, if δ is some logarithm of a rational number such that $0 < \delta < -\Lambda_{\text{cat}}(x||y)$, then Proposition 75 proves that $x \otimes s_\delta$ does not strongly noisy-trump y , so statement (B1) is false. This proves the lemma for the case $\Lambda_{\text{cat}}(x||y) < 0$.

In the remaining case $\Lambda_{\text{cat}}(x||y) \geq 0$, Proposition 49 implies that x noisy-trumps y . By definition, for every $\delta > 0$, there is a catalyst z such that $x \otimes s_\delta \otimes z \xrightarrow{\text{noisy}} y \otimes z$. According to Definition 70, this means that $x \otimes s_\delta$ strongly noisy-trumps y . This shows that (B1) is true in this case, and completes the proof of the lemma. ■

This lemma shows that there is no third alternative definition of noisy trumping. A conceivable alternative definition would involve the possibility to consume an arbitrarily small amount of nonuniformity (i.e. s_δ for arbitrarily small $\delta > 0$) and at the same time allowing for non-perfect state production, i.e. conversion to arbitrary non-perfect accuracy. Formally, this would be (B1); the lemma shows, however, that this only reproduces the notion of noisy-trumping from the main text. This result is not surprising due to the ‘‘closedness’’ property of noisy-trumping which was proven in Corollary 47.

Appendix C: Approximate distillation

The goal of this section is to give an expression for the maximal nonuniformity of any sharp state that can be extracted from some given state x , as announced in Eq. (6.5) in the main text. The result in Lemma 59 was achievable, but not optimal. To this end, we need a few lemmas.

Lemma 77 *Let \mathcal{D} be any contractive metric. Consider a bipartite classical system AB of finite dimension $d_{AB} = d_A d_B$. Let x^{AB} be any state on AB , with marginal x^A on A , and y^A any other state on A . Then there exists an extension y^{AB} of y^A (i.e. a state that has y^A as its marginal) with the property that*

$$\mathcal{D}(y^{AB}, x^{AB}) = \mathcal{D}(y^A, x^A).$$

Remark. In the quantum case, this statement is still true for the purified distance \mathcal{D}_p (Tomamichel, 2012), but the proof does not carry over to all contractive metrics on quantum states.

Proof. We can think of A as describing a random variable a that takes values in the discrete sample space $\{1, \dots, d_A\}$, and similarly for B . Then we use the shorthand notation

$$x_{i,j}^{AB} := \text{Prob}(a = i, b = j),$$

and we get $x_i^A = \sum_j x_{i,j}^{AB}$. This allows us to define the conditional probability distribution

$$x_{j|i}^{AB} := \begin{cases} x_{i,j}^{AB}/x_i^A & \text{if } x_i^A \neq 0, \\ \delta_{i,j} & \text{otherwise,} \end{cases}$$

where $\delta_{i,j} = 1$ if $i = j$ and 0 otherwise. We obtain

$$x_{i,j}^{AB} = x_{j|i}^{AB} \cdot x_i^A. \quad (\text{C1})$$

Moreover, since $x_{j|i}^{AB} \geq 0$ and

$$\sum_j x_{j|i}^{AB} = 1 \text{ for all } i,$$

the conditional probability $x_{j|i}^{AB}$ defines a stochastic matrix, i.e. a classical channel N from the probability distributions on A to those on AB , such that (C1) becomes $x^{AB} = N(x^A)$. Define $y^{AB} := N(y^A)$, then

$$\sum_j y_{i,j}^{AB} = \sum_j x_{j|i}^{AB} y_i^A = y_i^A,$$

such that y^{AB} is indeed an extension of y^A . Due to contractivity of \mathcal{D} , we obtain

$$\mathcal{D}(y^{AB}, x^{AB}) = \mathcal{D}(N(y^A), N(x^A)) \leq \mathcal{D}(y^A, x^A);$$

on the other hand, marginalization is also a channel, which proves the converse inequality. ■

This extension property implies an inequality for the max-entropy:

Lemma 78 *If x^{AB} is any bipartite state with marginal x^A , then*

$$H_0^\varepsilon(x^{AB}) \leq H_0^\varepsilon(x^A) + \log d_B$$

for every $\varepsilon \geq 0$.

Proof. First we prove the statement for $\varepsilon = 0$. Suppose that i is such that $x_i^A = 0$, then

$$x_i^A = 0 = \sum_j x_{i,j}^{AB} \Rightarrow x_{i,j}^{AB} = 0 \text{ for all } j.$$

Counting the zero-probability events, this implies that

$$\#\{(i, j) : x_{i,j}^{AB} = 0\} \geq d_B \cdot \#\{i : x_i^A = 0\}.$$

Thus,

$$\begin{aligned} 2^{H_0(x^{AB})} &= \#\{(i, j) : x_{i,j}^{AB} \neq 0\} \\ &= d_{AB} - \#\{(i, j) : x_{i,j}^{AB} = 0\} \\ &\leq d_A d_B - d_B \cdot \#\{i : x_i^A = 0\} \\ &= d_B \cdot \#\{i : x_i^A \neq 0\} \\ &= d_B \cdot 2^{H_0(x^A)}. \end{aligned}$$

Now we turn to the case $\varepsilon > 0$. Let \tilde{x}^A be any state that achieves the optimization in the definition of $H_0^\varepsilon(x^A)$, i.e. $\mathcal{D}(\tilde{x}^A, x^A) \leq \varepsilon$ and $H_0^\varepsilon(x^A) = H_0(\tilde{x}^A)$. According to Lemma 77, there is an extension \tilde{x}^{AB} of \tilde{x}^A such that $\mathcal{D}(\tilde{x}^{AB}, x^{AB}) = \mathcal{D}(\tilde{x}^A, x^A) \leq \varepsilon$. Thus

$$\begin{aligned} H_0^\varepsilon(x^{AB}) &\leq H_0(\tilde{x}^{AB}) \leq H_0(\tilde{x}^A) + \log d_B \\ &= H_0^\varepsilon(x^A) + \log d_B. \end{aligned}$$

This proves the claim. ■

Now we have collected enough information to prove our main result on the distillable purity.

Lemma 79 *Any distribution x can be converted by noisy operations into another distribution which is ε -close to a sharp state of nonuniformity*

$$I_0^\varepsilon(x \otimes m^{(d)}),$$

where $m^{(d)}$ is the uniform state in \mathbb{R}^d . This expression is increasing in d . Furthermore, the maximal nonuniformity of any sharp state that can be extracted from x is given by

$$\lim_{d \rightarrow \infty} I_0^\varepsilon(x \otimes m^{(d)}). \quad (\text{C2})$$

Proof. Clearly, the first statement is a simple consequence of Lemma 62: before extracting nonuniformity, we can add on a uniform state of any dimension we like. Lemma 78 implies that

$$\begin{aligned} I_0^\varepsilon(x \otimes m^{(d)}) &= \log(d_x \cdot d) - H_0^\varepsilon(x \otimes m^{(d)}) \\ &\geq \log d_x + \log d - H_0^\varepsilon(x) - \log d \\ &= I_0^\varepsilon(x), \end{aligned}$$

so that adding on a uniform state does not decrease the value of I_0^ε . Hence $I_0^\varepsilon(x \otimes m^{(d)})$ is increasing in d , and the expression in (C2) exists as a real number or plus infinity. For the time being, denote this expression by $J_0^\varepsilon(x)$. Generalizing the previous calculation, we see that Lemma 78 also implies that

$$I_0^\varepsilon(x^{AB}) \geq I_0^\varepsilon(x^A),$$

and so also $J_0^\varepsilon(x^{AB}) \geq J_0^\varepsilon(x^A)$. In other words, J_0^ε is non-increasing under marginalization. Furthermore, by construction, $J_0^\varepsilon(x \otimes m^{(d)}) = J_0^\varepsilon(x)$; that is, adding on maximally mixed states leaves J_0^ε invariant. Finally, note that

$$J_0^\varepsilon(\Pi(x)) = J_0^\varepsilon(x)$$

for any representation Π of a permutation; this property is inherited from I_0^ε and follows from the permutation-invariance of any contractive metric. These three properties together show that J_0^ε is a nonuniformity monotone.

Now suppose that $x \xrightarrow{\text{noisy}} y$ such that $\mathcal{D}(y, s_I) \leq \varepsilon$ for the sharp state s_I with nonuniformity I . Then

$$H_0^\varepsilon(y) \leq H_0(s_I) = \log d_u,$$

if $I = \log d/d_u$ for $d = d_y$. It follows that $I_0^\varepsilon(y) \geq I$, and since J_0^ε is a nonuniformity monotone,

$$J_0^\varepsilon(x) \geq J_0^\varepsilon(y) \geq I_0^\varepsilon(y) \geq I.$$

This shows that the expression in (C2) is optimal. ■

References

- Åberg, J., 2013, Nat. Commun. **4**, 1925.
 Allahverdyan, A. E., R. Balian, and T. M. Nieuwenhuizen, 2004, Journal of Modern Optics **51**(16-18), 2703, URL <http://dx.doi.org/10.1080/09500340408231829>.
 Amato, V., 1968, *Metodologia statistica strutturale* (F. Cacucci).
 Arnold, B. C., 1987, *Majorization and the Lorenz Order: A Brief Introduction* (Springer-Verlag).
 Audenaert, K. M. R., 2007, J. Phys. A: Math. Theor. **40**, 8127.
 Bartlett, S. D., T. Rudolph, and R. W. Spekkens, 2007, Rev. Mod. Phys. **79**, 555.
 Bennett, C. H., 1973, IBM J. Res. Dev. **17**, 525.
 Bennett, C. H., 1982, Int. J. Theor. Phys. **21**(12), 905.
 Bhatia, R., 1997, *Matrix analysis*, volume 169 (Springer).
 Birkhoff, G., 1946, Revi. Univ. Nac. Tucuman, ser A **5**, 147.
 Brandão, F. G., M. Horodecki, N. H. Y. Ng, J. Oppenheim, and S. Wehner, 2015, Proc. Natl. Acad. Sci. **112**, 3275.
 Brandão, F. G., M. Horodecki, J. Oppenheim, J. Renes, and R. W. Spekkens, 2013, Phys. Rev. Lett. **111**, 250404.
 Burg, J. P., 1967, in *Proceedings 37th Annual Meeting of the Society of Exploration Geophysicists*.
 Chefles, A., 2002, Phys. Rev. A **65**(5), 052314.
 Clausius, R., 1867, *The mechanical theory of heat: With its applications to the steam-engine and to the physical properties of bodies* (J. van Voorst).

- Coecke, B., T. Fritz, and R. W. Spekkens, 2014, arXiv:1409.5531 .
- Cover, T. M., and J. A. Thomas, 2012, *Elements of Information Theory* (John Wiley & Sons).
- Csiszár, I., 1967, *Studia Sci. Math. Hungar.* **2**, 299.
- Dahlsten, O. C., R. Renner, E. Rieper, and V. Vedral, 2009, arXiv:0908.0424 .
- van Dam, W., and P. Hayden, 2003, *Phys. Rev. A* **67**, 060302.
- Datta, N., 2009, *IEEE T. Inform. Theory* **55**(6), 2816.
- Datta, N., 2013, <http://pirsa.org/13060003/>.
- Del Rio, L., J. Åberg, R. Renner, O. Dahlsten, and V. Vedral, 2011, *Nature* **474**(7349), 61.
- Egloff, D., O. C. O. Dahlsten, R. Renner, and V. Vedral, 2012, arXiv:1207.0434 .
- van Erven, T., and P. Harremoës, 2010, in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on (IEEE)*, pp. 1335–1339.
- Faist, P., F. Dupuis, J. Oppenheim, and R. Renner, 2012, arXiv:1211.1037v1 .
- Fannes, M., 1973, *Commun. Math. Phys.* **31**, 291.
- Gini, C., 1912, *Studi Economico-Giuricici della R.*
- Gorban, A. N., P. A. Gorban, and G. Judge, 2010, *Entropy* **12**(5), 1145.
- Gour, G., 2005, *Phys. Rev. A* **71**, 012318.
- Gour, G., and R. W. Spekkens, 2008, *New Journal of Physics* **10**(3), 033023.
- Haagerup, U., and M. Musat, 2011, *Commun. Math. Phys.* **303**(2), 555.
- Hardy, G. H., J. E. Littlewood, and G. Polya, 1952, *Inequalities* (Cambridge university press).
- Horn, R. A., and C. R. Johnson, 2012, *Matrix Analysis* (Cambridge University Press).
- Horodecki, M., K. Horodecki, P. Horodecki, R. Horodecki, J. Oppenheim, A. Sen(De), and U. Sen, 2003a, *Phys. Rev. Lett.* **90**, 100402.
- Horodecki, M., P. Horodecki, and J. Oppenheim, 2003b, *Phys. Rev. A* **67**, 062104.
- Horodecki, M., and J. Oppenheim, 2013, *Nat. Commun.* **4**, 1.
- Horodecki, R., P. Horodecki, M. Horodecki, and K. Horodecki, 2009, *Rev. Mod. Phys.* **81**(2), 865.
- Janzing, D., P. Wocjan, R. Zeier, R. Geiss, and T. Beth, 2000, *Int. J. Theor. Phys.* **39**(12), 2717.
- Jaynes, E. T., 1957a, *Phys. Rev.* **106**, 620.
- Jaynes, E. T., 1957b, *Phys. Rev.* **108**, 171.
- Joe, H., 1990, *J. Math. Anal. Appl.* **148**(2), 287.
- Jonathan, D., and M. B. Plenio, 1999, *Phys. Rev. Lett.* **83**, 3566.
- Kakwani, N. C., 1980, *Income Inequality and Poverty: Methods of Estimation and Policy Applications* (Oxford University Press New York).
- Klimesh, M., 2007, arXiv:0709.3680 .
- Kullback, S., and R. A. Leibler, 1951, *The Annals of Mathematical Statistics* **22**(1), 79.
- Landauer, R., 1961, *IBM J. Res. Dev.* **5**, 183.
- Leff, H., and A. F. Rex (eds.), 2003, *Maxwell's Demons 2* (IOP Publishing).
- Lewenstein, M., and A. Sanpera, 1998, *Phys. Rev. Lett.* **80**(11), 2261.
- Lieb, E. H., and J. Yngvason, 1999, *Physics Reports* **310**(1), 1, ISSN 0370-1573, URL <http://www.sciencedirect.com/science/article/pii/S0370157398000829>.
- Linden, N., S. Popescu, and P. Skrzypczyk, 2010, *Phys. Rev. Lett.* **105**(13), 130401.
- Lorenz, M. O., 1905, *Publ. Am. Stat. Assoc.* **9**(70), 209.
- Marshall, A. W., I. Olkin, and B. C. Arnold, 2010, *Inequalities: theory of majorization and its applications* (Springer).
- Maruyama, K., F. Nori, and V. Vedral, 2009, *Rev. Mod. Phys.* **81**(1), 1.
- Marvian, I., and R. W. Spekkens, 2013, *New J. Phys.* **15**, 033001.
- Mendl, C. B., and M. M. Wolf, 2009, *Commun. Math. Phys.* **289**(3), 1057.
- Morimoto, T., 1963, *J. Phys. Soc. Jap* **12**, 328.
- Narang, G., and Arvind, 2007, *Phys. Rev. A* **75**, 032305.
- Nielsen, M., 2002, lecture notes (available online), URL <http://michaelnielsen.org/blog/talks/2002/maj/book.ps>.
- Nielsen, M. A., 1999, *Phys. Rev. Lett.* **83**(2), 436.
- Nielsen, M. A., and I. L. Chuang, 2010, *Quantum Computation and Quantum Information* (Cambridge University Press).
- Nielsen, M. A., and G. Vidal, 2001, *Quantum Inf. Comput.* **1**(1), 76.
- Rao, Y. V. C., 1997, *Chemical engineering thermodynamics* (Universities Press).
- Renner, R., 2008, *Int. J. Quantum Inf.* **6**(01), 1.
- Renner, R., and S. Wolf, 2004, in *ISIT Proceedings: International Symposium on Information Theory, 2004.* (IEEE), p. 233.
- Ruch, E., 1975a, *Theoretica Chimica Acta* **19**, 225.
- Ruch, E., 1975b, *Theoretica Chimica Acta* **38**, 167.
- Ruch, E., and B. Lesche, 1978, *J. Chem. Phys.* **69**(1), 393.
- Ruch, E., and A. Mead, 1976, *Theoretica Chimica Acta* **41**(2), 95.
- Ruch, E., R. Schraner, and T. H. Seligman, 1978, *The Journal of Chemical Physics* **69**, 386.
- Sanders, Y. R., and G. Gour, 2009, *Phys. Rev. A* **79**, 054302.
- Schutz, R. R., 1951, *Amer. Econ. Rev.* **41**, 107.
- Shor, P. W., 2010, *Structure of Unital Maps and the Asymptotic Quantum Birkhoff Conjecture*, presentation.
- Skrzypczyk, P., A. J. Short, and S. Popescu, 2013, arXiv:1302.2811 .
- Skrzypczyk, P., A. J. Short, and S. Popescu, 2014, *Nat. Comm.* **5**, 4185.
- Szilard, L., 1929, *Zeitschrift für Physik* **53**(11-12), 840.
- Terhal, B. M., I. L. Chuang, D. P. DiVincenzo, M. Grassl, and J. A. Smolin, 1999, *Phys. Rev. A* **60**(2), 881.
- Tomamichel, M., 2012, arXiv:1203.2142 .
- Tsallis, C., 1988, *J. Stat. Phys.* **52**, 479.
- Tsallis, C., 1998, *Phys. Rev. E* **58**, 1442.
- Turgut, S., 2007, *J. Phys. A: Math. Theor.* **40**(40), 12185.
- Uhlmann, A., 1970, *Rep. Math. Phys.* **1**(2), 147.
- Veinott, A. F., 1971, *Manage. Sci.* **17**(9), 547.
- Veitch, V., S. A. H. Mousavian, D. Gottesman, and J. Emerson, 2014, *New J. Phys.* **16**, 013009.
- Wallace, D., 2013, arXiv:1306.4907 .
- Zalka, C., and E. Rieffel, 2002, *J. Mat. Phys.* **43**, 4376.