Variance of Relative Surprisal as Single-Shot Quantifier

Paul Boes,[1,*] Nelly H.Y. Ng,[1,2,†] and Henrik Wilming[3,4,‡]

[1]*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Berlin 14195, Germany*

[2]*School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, 637371, Singapore*

[3]*Institute for Theoretical Physics, ETH Zurich, Zurich 8093, Switzerland*

[4]*Leibniz Universität Hannover, Appelstraße 2, Hannover 30167, Germany*

The variance of (relative) surprisal, also known as varentropy, so far mostly plays a role in information theory as quantifying the leading-order corrections to asymptotic independent and identically distributed (IID) limits. Here, we comprehensively study the use of it to derive single-shot results in (quantum) information theory. We show that it gives genuine sufficient and necessary conditions for approximate state transitions between pairs of quantum states in the single-shot setting, without the need for further optimization. We also clarify its relation to smoothed min and max entropies, and construct a monotone for resource theories using only the standard (relative) entropy and variance of (relative) surprisal. This immediately gives rise to enhanced lower bounds for entropy production in random processes. We establish certain properties of the variance of relative surprisal, which will be useful for further investigations, such as uniform continuity and upper bounds on the violation of subadditivity. Motivated by our results, we further derive a simple and physically appealing axiomatic single-shot characterization of (relative) entropy, which we believe to be of independent interest. We illustrate our results with several applications, ranging from interconvertibility of ergodic states, over Landauer erasure to a bound on the necessary dimension of the catalyst for catalytic state transitions and Boltzmann's *H* theorem.

I. INTRODUCTION

Many central results of quantum information theory are concerned with the manipulation of quantum systems in the so-called asymptotic independent identically distributed (IID) setting, in which one considers the limit of taking infinitely many independent and identically distributed copies of a quantum system [1–9]. While not always being physically realistic, this setting is convenient to work in because it allows for the application of standard concentration results from statistics and information theory [10,11]. Recent years have seen a lot of effort in studying more general settings, in which subsystems might be correlated with one another or the size of the

---
*pboes@zedat.fu-berlin.de

†nelly.hy.ng@gmail.com

‡henrik-physics@arrr.de

system is finite. Conceptually, the most extreme weakening of the IID setting is the *single-shot* setting, in which generally no assumptions about the size of the system or its correlations are made. This setting derives its name from the fact that it can be seen to describe a single iteration of a protocol, in contrast to the IID setting, which is concerned with infinitely many independent iterations. Several classic results [12–16] have been instrumental to the recent developments of majorization-based resource theories [8,17–23], which have found diverse applications in entanglement theory, thermodynamics, asymmetry, and many more central topics in information theory. By now there exists a detailed understanding of an intuitive trade-off between the above settings (and that we detail in later sections). On the one hand, the IID setting, due to its various assumptions, can usually be characterized by variants of the (quantum) relative entropy,

$$S(\rho \| \sigma) := \mathrm{tr}\{\rho[\log(\rho) - \log(\sigma)]\}.$$

On the other hand, majorizationlike conditions play a central role for single-shot transformations, even extending to the study of approximate transitions. Many such results are tight in the quasiclassical regime [24,25] (when we restrict to transformations between pairs of states that commute

with another). Nevertheless, the number of conditions that need to be checked increases linearly with the dimension of the states of interest, which makes a more systematic and simplified understanding in characterizing the possibility of state transitions difficult. Fortunately, the family of *smoothed entropies* have turned out to be a powerful tool to describe these constraints and operationally characterize a variety of single-shot tasks [22,26–39]. However, they typically require an optimization over the potentially high-dimensional state space.

In this work, we develop a complementary approach to the study of the single-shot setting in which "single-shot effects" are witnessed and quantified by a single quantity, the *variance of (relative) surprisal*,

$$V(\rho \| \sigma) := \mathrm{tr}\left[\rho \log^2\left(\frac{\rho}{\sigma}\right)\right] - S(\rho \| \sigma)^2, \qquad (1)$$

where $\log(\rho/\sigma) \equiv \log \rho - \log \sigma$ is the *relative surprisal* of one quantum state ρ with respect to another σ, and we use $\log \equiv \log_2$. In the following, we refer to this quantity simply as the relative variance. The relative variance has been shown to measure leading-order corrections to asymptotic results in (quantum) information theory [40–56], but here we show that its role extends to the genuine single-shot setting. We do this from two points of view: first, we show that the relative variance quantifies single-shot corrections to possible state transitions between pairs of quantum states. These imply that state transitions between pairs of states with low relative variance are essentially characterized by the relative entropy and hence do not exhibit strong single-shot effects. Simple examples of such states are ergodic states (see below).

Since single-shot effects often represent operational impediments (such as, for example, decreased ability to extract work in the context of thermodynamics), the above finding motivates the question whether there is, in some sense, a "cost" associated to obtaining states of low relative variance. We proceed to show that this is indeed the case: reducing the relative variance between a pair of states necessitates a proportional reduction of the relative entropy between those states. In the resource-theoretic setting (defined later), in which the relative entropy is known to itself be measure of a state's resourcefulness, this finding implies that increasing the operational value of a system in the sense of reducing its relative variance comes at the cost of reducing its value as measured by the relative entropy. Formally, the above trade-off result is an implication of a resource monotone that we construct. We formulate the trade-off between relative variance and relative entropy both for single systems as well as for the *marginal* changes of bipartite systems.

Overall, our findings motivate the relative variance as a simple measure of single-shot effects: the smaller the relative variance of pairs of states, the better their single-shot properties are described by the relative entropy, and

vice versa. In addition to the above results, we clarify the relation between the relative variance and the smoothed entropies, relate our single-shot results to the known asymptotic leading-order corrections mentioned above, and present a novel axiomatic characterization of the (relative) entropy from a single, physically motivated axiom. We believe the latter finding to be of independent interest. Apart from its applications in the context of resource theories, our work can also be seen as a thorough investigation of the mathematical properties of the variance of relative surprisal, such as uniform continuity bounds, corrections to subadditivity and the relation of the variance of relative surprisal to smoothed Rényi divergences and the theory of approximate majorization.

The remainder of this paper is structured as follows: our results are concerned with generic state transitions between pairs of states in the quasiclassical setting. However, following a formal introduction of the setup and terminology below, for the sake of clarity, in Sec. II we first provide an overview of our main results for the special case of state transitions under unital channels, which also corresponds to the resource theory of purity. The results are then shown in full generality and further discussed in Sec. III. Throughout, we focus on the formal results and provide applications mostly for illustration in boxes, leaving a more detailed study of applications and implications of our results to future work.

A. Setup and notation

Let S and S' be quantum systems represented on Hilbert spaces with fixed and finite respective dimensions d and d'. Further, let $\mathcal{D}(S)$ be the set of quantum states on S, and similarly for S'. Given two pairs of states $\rho, \sigma \in \mathcal{D}(S)$ and $\rho', \sigma' \in \mathcal{D}(S')$—sometimes referred to as dichotomies—we write $(\rho, \sigma) \succ (\rho', \sigma')$ if there exists a quantum channel $\mathcal{E} : \mathcal{D}(S) \to \mathcal{D}(S')$ such that $\mathcal{E}(\rho) = \rho'$ and $\mathcal{E}(\sigma) = \sigma'$. The preorder $\succ$ has been studied both classically and quantumly (e.g., Refs. [13,17,57,58]) and forms the backbone of many *resource theories*. In a resource theory a set of so-called *free states* $\mathcal{D}_\mathcal{F} \subseteq \mathcal{D}(S)$ is specified, together with a set of quantum channels $\mathcal{F}$ such that each of these channels maps free states into free states, $\mathcal{F}[\mathcal{D}_\mathcal{F}] \subseteq \mathcal{D}_\mathcal{F}$. These channels are therefore called *free*. This terminology originates from the idea that free states constitute a class of states that are easy to prepare in a given physical context, while free channels constitute physical operations that are easy to implement in this context (see Ref. [23] for a review on resource theories). One is then concerned with the special case of the ordering $\succ$ in which $\mathcal{E}$ is free. In the special case that there is only a single free state σ, this corresponds to the special case $\sigma = \sigma'$ and induces the preorder $\succ_\sigma$ on $\mathcal{D}(S)$, known as σ majorization, where $\rho \succ_\sigma \rho'$ is equivalent to $(\rho, \sigma) \succ (\rho', \sigma)$. As

such, the results presented below can naturally be applied to those resource theories that follow the above form, although we emphasize that they hold more generally as well.

Since we are often concerned with approximate state transitions, we further write $(\rho, \sigma) \succ_\epsilon (\rho', \sigma')$ whenever there exists a state ρ'_ϵ such that $(\rho, \sigma) \succ (\rho'_\epsilon, \sigma')$ and $D(\rho', \rho'_\epsilon) := \frac{1}{2} \|\rho' - \rho'_\epsilon\|_1 \leq \epsilon$.

Finally, the IID setting corresponds to the special case of dichotomies of the form $(\rho^{\otimes n}, \sigma^{\otimes n}) \succ (\rho'^{\otimes n}, \sigma'^{\otimes n})$ for some $n \in \mathbb{N}$. Here, one usually considers the asymptotic limit $n \to \infty$. The statement we made in the introduction, that in this limit the relative entropy characterizes possible state transitions, is based on the well-known fact that, for any two dichotomies (ρ, σ) and (ρ', σ') it holds that, for any $\epsilon > 0$, there exists a sufficiently large n such that $(\rho^{\otimes n}, \sigma^{\otimes n}) \succ_\epsilon (\rho'^{\otimes n}, \sigma'^{\otimes n})$ if and only if $S(\rho\|\sigma) > S(\rho'\|\sigma')$. We recover this statement as a special case from the results below.

II. OVERVIEW OF MAIN RESULTS

In this section, we provide an overview of our results for the special case $\succ_\mathbb{I}$, where

$$\mathbb{I} \equiv \mathbb{1}_d / d$$

is our shorthand for the maximally mixed state in d dimensions, whereas $\mathbb{1}_d$ is the d-dimensional identity operator. This choice of σ corresponds to the resource theory of purity, also known as the resource theory of stochastic nonequilibrium [59], and captures the essential insights from our results while being easier to state.

Channels that preserve the maximally mixed state are also called *unital channels* [60] and the ordering $\succ_\mathbb{I}$ is known as *majorization* [61]. Unital channels are often used to model random processes. For instance, a (strict) subset of unital channels are random unitary channels that describe the evolution of a system under a unitary operator that was drawn at random from some fixed distribution. As such, the resource theory of purity is concerned with describing the evolution and operational "value" of quantum states in the presence of random processes.

A. Variance of surprisal

The central quantity in this work is the relative variance, defined in Eq. (1). For $\sigma = \mathbb{I}$, the relative variance reduces to the *variance (of surprisal)*

$$V(\rho) := V(\rho\|\mathbb{I}) = \text{tr}[\rho \log^2(\rho)] - S(\rho)^2,$$

where $S(\rho) := \log(d) - S(\rho\|\mathbb{I}) = -\text{tr}[\rho \log(\rho)]$ is the von Neumann entropy [itself the mean of the *surprisal*, $-\log(\rho)$]. The variance of surprisal is also known as *information variance* or *varentropy*, and as *capacity of*

entanglement in the context of entanglement in many-body physics [62–64].

Operationally, the variance can be understood, for instance, as the variance of the length of a codeword in an optimal quantum source code. However, while, as mentioned in the introduction, it is well known to quantify the leading-order corrections to various (quantum) information-theoretic tasks in the asymptotic limit, its relevance in the single-shot setting has not yet, to the authors' knowledge, been thoroughly investigated (recently, some formal properties have, however, been developed in Ref. [65]). In this work, we study the (relative) variance and show that it in fact provides a useful measure of single-shot effects for approximate state transitions.

We begin by mentioning some properties that the variance of surprisal fulfills, and which we use throughout the paper:

1. Additivity under tensor products:

$$V(\rho_1 \otimes \rho_2) = V(\rho_1) + V(\rho_2).$$

2. Positivity: $V(\rho) \geq 0$.
3. Uniform continuity (Lemma 10):

$$|V(\rho) - V(\rho')|^2 \leq K \log^2(d) D(\rho, \rho')$$

   for a constant K.
4. Correction to subadditivity (Lemma 11):

$$V(\rho) \leq V(\rho_1) + V(\rho_2) + K' \log^2(d) f(I_\rho),$$

   for any bipartite state ρ with respective marginal states ρ_1, ρ_2 and mutual information I_ρ, with constant K' and $f(x) = \max\{\sqrt[4]{x}, x^2\}$.
5. $V(\rho) = 0$ if and only if all nonzero eigenvalues of ρ are the same. We call such states *flat states*. Examples include any pure state and the maximally mixed state.
6. For fixed dimension $d \geq 2$, the state $\hat{\rho}_d$ with maximal variance has the spectrum [66]

$$\text{spec}(\hat{\rho}_d) = \left(1 - r, \frac{r}{d-1}, \ldots, \frac{r}{d-1}\right), \quad (2)$$

   with r being the unique solution to

$$(1 - 2r) \ln\left(\frac{1-r}{r}(d-1)\right) = 2.$$

   We have $(1/4)\log^2(d-1) < V(\hat{\rho}_d) < (1/4)\log^2(d-1) + 1/\ln^2(2)$, and, in the limit of large d, $r \approx \frac{1}{2}$.

Properties 3 and 4 are original contributions of this work and are part of our main technical results.

B. Sufficient criteria for single-shot state transitions

It is well known that when considering the IID limit, approximate majorization reduces to an ordering with respect to the von Neumann entropy. More precisely, given two states ρ and ρ', then $S(\rho') > S(\rho)$ implies that for any $\epsilon > 0$ there exists a number $N_\epsilon \in \mathbb{N}$ such that

$$\rho^{\otimes n} \succ_{\mathbb{I},\epsilon} \rho'^{\otimes n} \quad \forall\, n \geq N_\epsilon.$$

However, this is not the case in a single-shot setting. Here, the question whether $\rho \succ_{\mathbb{I}} \rho'$ in full generality depends on $d - 1$ independent constraints on the spectra of ρ and ρ'. This makes dealing with exact single-shot state transitions considerably more difficult.

Our first result shows that for approximate state transitions, there nevertheless exist simple sufficient conditions at the single-shot level that also involve the von Neumann entropy, but with a correction quantified by the variance.

Result 1. (Sufficient conditions for approximate state transition). *Let ρ, ρ' be two states on S and $1 > \epsilon > 0$. If*

$$S(\rho') - \sqrt{V(\rho')(2\epsilon^{-1} - 1)} \geq S(\rho) + \sqrt{V(\rho)(2\epsilon^{-1} - 1)},$$

then $\rho \succ_{\mathbb{I},\epsilon} \rho'$.

We emphasize that this result is a fully single-shot result. It shows that the variance quantifies the single-shot deviation from the above IID case. A convenient reformulation of Result 1 is as follows: let ρ, ρ' be two states on S with

$$S(\rho') - S(\rho) = \delta > 0.$$

Solving for ϵ in Result 1, we see that $\rho \succ_{\mathbb{I},\epsilon} \rho'$, where

$$\epsilon \leq \frac{2}{\delta^2}\left[\sqrt{V(\rho)} + \sqrt{V(\rho')}\right]^2,$$

can be achieved. An appealing feature of this result is that it does not require any optimization, as is typically present in results relying on smoothed entropies (compare, for example, Result 1 and its generalization Theorem 8 to the results in Ref. [58]). When applied to state transitions under unital channels in the IID limit, Result 1 straightforwardly produces finite-size corrections towards asymptotic interconvertibility: given two states ρ and ρ' with $S(\rho') > S(\rho)$, it implies that $\rho^{\otimes n} \succ_{\mathbb{I},\epsilon} (\rho')^{\otimes n}$ with

$$\epsilon \leq \frac{2[\sqrt{V(\rho)} + \sqrt{V(\rho')}]^2}{n[S(\rho') - S(\rho)]^2}, \tag{3}$$

which vanishes in the limit $n \to \infty$. For finitely many copies of the two states, the variances of initial and final states bound the achievable precision.

In Appendix F we provide a more detailed analysis of the IID case, where we use Result 1 (and its generalization to σ majorization) to study convertibility between sequences of n IID states for large but finite n and with an error ϵ_n such that $n\epsilon_n \to \infty$, but possibly $\epsilon_n \to 0$. This can be seen as a simple form of moderate-deviation analysis, and in particular we recover the "resonance" phenomenon reported in Refs. [55,56], namely that second-order corrections vanish when

$$\frac{V(\rho)/S(\rho)}{V(\rho')/S(\rho')} = 1,$$

with the proof being as simple as solving a quadratic equation.

Result 1 implies that state transitions between pairs of initial and final states with low variance are essentially characterized by the entropy. As an application, in Box 1, we prove a simpler version of a recent result on the macroscopic interconvertibility of ergodic states under thermal operations [67,68]. Finally, let us note that Eq. (3) is in general not very tight, since we know from Hoeffding-type bounds that, in the IID limit, the amount of probability outside of the typical window (which directly contributes to the error) scales as $\epsilon \propto \exp(-n)$. Nevertheless, the absence of any trailing terms makes it simple to evaluate, especially in single-shot scenarios.

Box 1. Interconvertibility of ergodic states

As an application of Result (1), we discuss the interconvertibility of *ergodic states* in the unital setting. Informally speaking, ergodic states are states on an infinite chain of identical, finite-dimensional Hilbert spaces of dimension d, enumerated by $\mathbb{Z}$ and called sites below, which have the property that correlations between observables located at distance sites converge to zero as their distance is increased. See Ref. [67] for a detailed description of ergodic states. Importantly, states are in general correlated, examples being ground states of gapped, local Hamiltonians or thermal states of many-body systems away from the critical temperature. Nevertheless, if ρ_n denotes the density matrix of n consecutive sites of the chain with entropy $S_n = S(\rho_n)$, the (quantum) Shannon-MacMillan-Breimann theorem shows that for arbitrarily small $\epsilon > 0$ and sufficiently large n, we can find an approximation ρ_n^ϵ of ρ_n with the property that each eigenvalue p_j of ρ_n^ϵ fulfills [69]

$$|-\log(p_j) - S_n| \leq 2n\epsilon$$

and $D(\rho_n, \rho_n^\epsilon) \leq \epsilon$. Thus, the variance fulfills

$$
\begin{aligned}
V(\rho_n^\epsilon) &= \sum_j p_j \left[\log(1/p_j) - S_n^\epsilon \right]^2 \\
&= \sum_j p_j \left[\log(1/p_j) - S_n \right]^2 - (S_n - S_n^\epsilon)^2 \\
&\leq 4n^2 \epsilon^2,
\end{aligned}
$$

where $S_n^\epsilon = S(\rho_n^\epsilon)$. By uniform continuity of the variance, Lemma 10, we therefore have

$$
V(\rho_n) \leq V(\rho_n^\epsilon) + Kn^2\sqrt{\epsilon} \leq \tilde{K}^2 n^2 \sqrt{\epsilon},
$$

with $\tilde{K}$ some constant and where we use $\epsilon^2 \leq \sqrt{\epsilon}$. Result 1 now tells us that if we have two ergodic states with entropies $S_n = sn$ and $S_n' = s'n$ such that $s < s'$, then for any $\epsilon > 0$ and sufficiently large n, we can convert ρ_n to ρ_n' using a unital channel with error at most

$$
2 \left[\frac{\sqrt{V(\rho_n)} + \sqrt{V(\rho_n')}}{(s'-s)n} \right]^2 \leq \frac{16\tilde{K}^2}{(s'-s)^2} \sqrt{\epsilon},
$$

which can be made arbitrarily small.

C. Relation to smoothed min and max entropies

As mentioned in the introduction, smoothed generalized entropies are often used to describe single-shot processes. For instance, the continuous family of *Rényi entropies* has been found to characterize possible single-shot transitions in the semiclassical setting [37,70]. Among those entropies, the smoothed *min* and *max* entropies are of particular prominence, since they enjoy clear operational meanings in various information-processing tasks such as randomness extraction or data compression (see, e.g., Refs. [26,30]) and, in a sense, quantify complementary single-shot properties of quantum states. These quantities, the precise definition of which is given in Sec. III C, can differ significantly from the von Neumann entropy for arbitrary states. However, the following result shows that one can bound this difference by the variance.

Result 2. (Bounds on smoothed min and max entropies). *Let $1 > \epsilon > 0$ and let ρ be a state on S. Then,*

$$
S_{\max}^\epsilon(\rho) - S(\rho) \leq \sqrt{(\epsilon^{-1} - 1)V(\rho)},
$$
$$
S(\rho) - S_{\min}^\epsilon(\rho) \leq \sqrt{(\epsilon^{-1} - 1)V(\rho)}.
$$

As a straightforward corollary, Result 2 implies that for any $1 > \epsilon > 0$ and any state ρ,

$$
S_{\max}^\epsilon(\rho) - S_{\min}^\epsilon(\rho) \leq 2\sqrt{(\epsilon^{-1} - 1)V(\rho)}. \tag{4}
$$

Result 2 has the appealing feature of providing an upper bound that factorizes into the variance and a function of the smoothing parameter ϵ. In analogy with Result 1, Result 2 shows that for states with small variance, finite-size corrections (e.g., to coding rates) are less pronounced, making these states ideal candidates for information encoding and transmission. This makes intuitive sense if we recall that the spectrum of states with zero variance is uniform over its support. Within the class of states with the same entropy, these *flat* states are therefore maximally "compressed" and "random." As a side note, we observe that various models for "batteries" used in single-shot quantum thermodynamics restrict to battery states with zero or very low relative entropy variance. This choice can conveniently be interpreted in terms of Eq. (4), since $S_{\min}^\epsilon$ and $S_{\max}^\epsilon$ quantify the amount of single-shot work of creation and extractable work, respectively, for nonequilibrium states [19,37]. Restricting to low-variance battery states therefore ensures that the process of storing and extracting work from a battery involves a minimal dissipation of heat.

D. Decrease of variance

The previous two results establish the variance as a measure of single-shot effects, bounding the extent of such effects in the context of approximate state transitions and operational tasks such as data compression or randomness extraction. In particular, they imply that the manipulation of states with low variance produces less overhead due to finite-size effects. Therefore, state transitions between states of low variance can be operationally advantageous. This motivates the question whether there exists a resource-theoretic "cost" associated to decreasing the variance of a state. We show that this is indeed the case: under unital channels, decreasing the variance lower bounds *entropy production*.

The statement above, formulated in Result 3 is a consequence of a new resource monotone that we derive.

Definition 1: (Resource monotone). Let $\mathcal{D}$ be the set of all quantum states. A function $f : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ is called a *(resource) monotone* if for any dichotomies satisfying $(\rho, \sigma) \succ (\rho', \sigma')$, we also have $f(\rho, \sigma) \geq f(\rho', \sigma')$.

Nonincreasing resource monotones with respect to majorization are called *Schur convex* and nondecreasing ones *Schur concave*. Resource monotones are an important tool in the study of resource theories. For example, for a Schur convex function f, $f(\rho') > f(\rho)$ suffices to conclude that $\rho \nsucc_\mathbb{I} \rho'$ but the former is often far easier

to check than the latter. An example of a Schur convex function is the purity $\text{tr}(\rho^2)$ of a state ρ, while the entropy $S(\rho)$ is Schur concave. The variance itself is evidently not monotone, which might partially explain why it has so far not been studied resource theoretically. However, the following lemma shows that the variance and entropy jointly give rise to a monotone.

Lemma 2. (*Schur concavity of M*). *The function*

$$M(\rho) := V(\rho) + \left(\frac{1}{\ln(2)} + S(\rho) \right)^2,$$

is Schur concave.

By Schur concavity, we have $0 \leq M(\rho) \leq [1/\ln(2) + \log(d)]^2$. Notably, unlike many commonly used monotones, M is not additive with respect to product states. Figure 1 compares the regions of increasing M and entropy compared to the majorization ordering for two initial states in $d = 3$ and fixed eigenbasis, illustrating that for some states M provides strictly stronger necessary conditions for the majorization ordering than the entropy S.

By means of Lemma 2, we can derive the following bound on entropy production, which is our third main result and proven in the more general statement of Corollary 13.

Result 3. (Lower bound on entropy production). *Let ρ, ρ' be a pair of states on S. If $\rho \succ_{\mathbb{1}} \rho'$, then*

$$S(\rho') - S(\rho) \geq \frac{V(\rho) - V(\rho')}{2\sqrt{M(\rho)}} \geq \frac{V(\rho) - V(\rho')}{2[1/\ln(2) + \log(d)]}. \tag{5}$$

Result 3 shows that the decrease of variance under unital channels can only come at the cost of increasing the system's entropy. It also complements the previous results, in that it shows that states with positive variance *necessarily* exhibit finite-size effects: the entropy fails to characterize single-shot transitions, as witnessed by the variance. Still, the IID limit is consistent with Eq. (5), since in the asymptotic limit the lhs grows linearly, while the other terms remain constant to leading order. At the same time, there exist sequences of state transitions for which the constraint imposed by Eq. (5) remains nontrivial in the limit of large system size. An example are transitions from the state $\hat{\rho}_d$, as defined by Eq. (2), to any state of constant variance, in the limit of growing d. As a simple application, in Box 2, we apply Result 3 to the task of erasure and find corrections to Landauer's principle that are quantified by the variance.
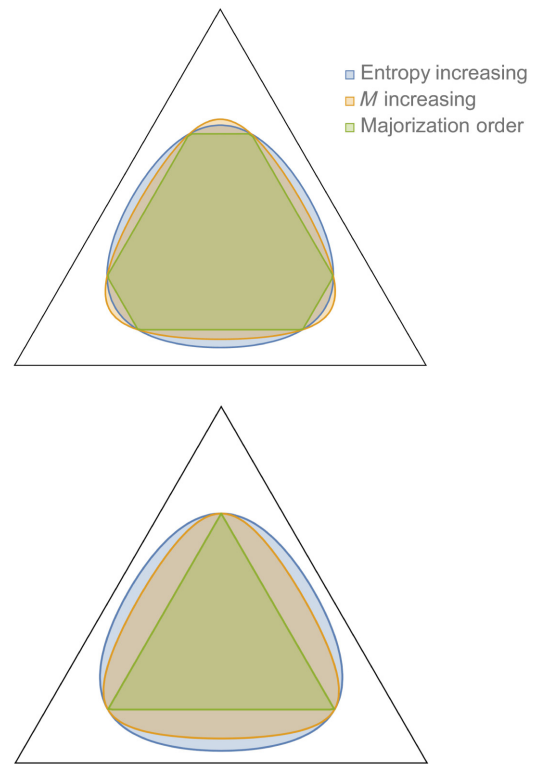

FIG. 1. Region plots of increasing M and entropy S for two initial states [top, $p = (0.65, 0.25, 0.1)$; bottom, $p = (0.7, 0.15, 0.15)$] in the simplex of states $\rho = \sum_{i=1}^{3} p_i |i\rangle\langle i|$, i.e., for $d = 3$ and fixed eigenbasis. The sets of states are shown in orange and blue with increasing M and S, respectively, while green shows the set of states majorized by the initial state. While in general neither the red or blue region contains the other (top), for some states M provides strictly stronger necessary conditions to rule out majorization (bottom).

Box 2. Finite-size corrections to Landauer's principle

Landauer's principle states that the erasure of information physically requires the dissipation of entropy, which consumes work, converting it into heat [71,72]. A simple resource-theoretical model of an erasure process consists of a unital channel acting on a system S whose state is to be erased (i.e., mapped to a fixed pure state $|\psi\rangle$) together with an *information battery* B that acts as a source of purity. In the simplest setting, B is an n-qubit system, with each qubit being either in a pure state $|0\rangle$ or maximally mixed. We define the *finite-size work cost* [in units of $k_B \ln(2)$] of erasing an initial state ρ as the size of the smallest information battery that allows for an erasure of ρ, i.e., the smallest integer n such

that

$$\rho \otimes |0\rangle\langle 0|^{\otimes n} \succ_{\mathbb{I}} |\psi\rangle\langle\psi| \otimes (\mathbb{1}_2/2)^{\otimes n}.$$

The usual formulation of Landauer's bound, $n \geq S(\rho)$, is then a simple consequence of the monotonicity of the entropy. However, applying Result 3 yields,

$$n \geq S(\rho) + \frac{V(\rho)}{2\sqrt{M(\rho)}}, \qquad (6)$$

which provides corrections to Landauer's bound that increase with the initial variance of ρ. Indeed, we note that Eq. (6) remains true even when the battery size is not constrained, i.e., one can allow the initial and final battery states to contain arbitrarily large "reservoirs" of pure and maximally mixed qubits, that is, states of the form $|0\rangle\langle 0|^{\otimes\lambda_1} \otimes (\mathbb{1}_2/2)^{\otimes\lambda_2}$ and $|0\rangle\langle 0|^{\otimes(\lambda_1-n)} \otimes (\mathbb{1}_2/2)^{\otimes(\lambda_2+n)}$ for arbitrarily large λ_1, λ_2. At the same time, the correction term in Eq. (6) can easily be made to vanish in the presence of a bystander system whose state is returned unchanged and uncorrelated from S and B. Such systems are technically known as trumping catalysts (see Box 3) and hence the above bound is not robust to this simple extension of the setting.

E. Bounds on marginal entropy production

In the previous subsection, it was shown that a decrease in the variance lower bounds entropy production under unital channels. However, not all quantum channels are unital channels and hence it is natural to ask whether a similar result exists for a more general class of channels. It is clear that Result 3 cannot be generalized for *all* channels: for instance, the channel that maps every input state to a pure state reduces both entropy and variance. However, using the properties of the variance and the previous results, we can formulate an analogous lower bound for arbitrary quantum channels, by considering their effect on the *environment*. To state those results, we first note the well-known fact that every quantum channel on a quantum system S can be understood as the local effect of a unital channel acting on S together with an environment E. More formally, any channel $\mathcal{E}$ from S to itself can be written as

$$\mathcal{E}(\cdot) = \mathrm{tr}_E[\mathcal{U}(\cdot \otimes \rho_E)]$$

for some initial state ρ_E of the environment and unital channel $\mathcal{U}$ on the joint system SE. We call the pair $(\mathcal{U}, \rho_E)$ a *dilation* of $\mathcal{E}$. By the Stinespring dilation theorem, one can always choose ρ_E to be pure and $\mathcal{U}$ to be a unitary

channel for sufficiently large environment dimension d_E, but here, in the context of the resource theory of purity, we use the above more general representation.

We next show that one can extend Result 3 to *local* changes of variance and entropy. Let $(\mathcal{U}, \rho_E)$ be a dilation of a given quantum channel $\mathcal{E}$. For an initial ρ_S on S, let $\rho'_S = \mathcal{E}(\rho_S)$ denote the final state on S and let $\rho'_E = \mathrm{tr}_S[\mathcal{U}(\rho_S \otimes \rho_E)]$ denote the final state on E. Moreover, denote as

$$\Delta S_S = S(\rho_S) - S(\rho'_S),$$
$$\Delta V_S = V(\rho_S) - V(\rho'_S),$$

the changes of entropy and variance on S, respectively, and similarly for the environment. Finally, let

$$I_{S:E} = S[\mathcal{U}(\rho_S \otimes \rho_E)\|\rho'_S \otimes \rho'_E] \qquad (7)$$

denote the mutual information between S and E after the application of $\mathcal{U}$. We then have the following.

Result 4. (Lower bound on *marginal* entropy production). *Given a quantum channel $\mathcal{E}$ from S to itself, let $(\mathcal{U}, \rho_E)$ be any dilation of $\mathcal{E}$ where $\mathcal{U}$ is a unital map, and denote $d_{SE} = d_S \cdot d_E$. Then, for any $\rho_S \in \mathcal{D}(S)$,*

$$-\Delta S_S - \Delta S_E \geq \frac{\Delta V_S + \Delta V_E - K' \log(d_{SE})^2 f\left(I_{S:E}\right)}{2\sqrt{M(\rho_S \otimes \rho_E)}}, \tag{8}$$

where K' is a constant independent of d or ρ and $f(x) := \max\{\sqrt[4]{x}, x^2\}$.

This result follows straightforwardly from combining Result 3 with Property 4 of the variance (or more precisely, Lemma 11) together with the subadditivity of the von Neumann entropy. It is particularly interesting from a resource-theoretic point of view, since in such theories, the environment E often explicitly models a particular kind of physical system, such as a thermal bath, a clock, a battery, or a catalyst. For instance, we may then consider the setting of Landauer erasure described in Box 2, with the additional requirement that $I_{S:E} = 0$. In Box 3, we also apply Result 4 to gain insight into catalytic processes, in particular, to derive bounds on the dimension of the catalyst required for certain processes.

Box 3. Bound on catalyst dimension for state transitions

It is well known that the set of possible state transitions in a resource theory can be enlarged with the help of catalysts, that is, auxiliary systems whose

---

local state remains unchanged in a process. In terms of the notation established in the main text and given two quantum states $\rho, \rho' \in \mathcal{D}(S)$, we write $\rho \succ_C \rho'$ if there exists a quantum channel $\mathcal{E}$ with dilation $(\mathcal{U}, \rho_E)$ such that $\rho'_E = \rho_E$, that is, if the local state of the environment remains unchanged. Moreover, we write $\rho \succ_T \rho'$ if the dilation can be chosen such that $I_{S:E} = 0$, where the catalyst not only remains locally unchanged, but is also returned uncorrelated from S. This relation is known as *trumping* [73–75]. Clearly,

$$\rho \succ_{\mathbb{I}} \rho' \Rightarrow \rho \succ_T \rho' \Rightarrow \rho \succ_C \rho',$$

while the converse relations in general do not hold. As such, catalysts enable previously impossible state transitions.

Indeed, recently it was shown that if ρ and ρ' are two full-rank states, then $\rho \succ_C \rho'$ is equivalent to $S(\rho') > S(\rho)$ [76], a result that has found applications in the context of fluctuation theorems in (quantum) thermodynamics [77] and can be further strengthened to the case that $\mathcal{U}$ is unitary [78,79]. What these results are silent about, however, is the required size of the catalyst. Here, we apply Result 4 to show that transitions between states with similar entropy that decrease the variance can only be realized by means of a catalyst with very large dimension. In particular, consider any state transition $\rho \succ_C \rho'$ between full-rank states such that $0 \leq S(\rho') - S(\rho) \leq \delta \leq 1$. Then, Result 4 implies that

$$\Delta V_S \leq \tilde{K} \log(d_{SE})^2 \sqrt[4]{\delta}.$$

This follows from the fact that $\Delta V_E = \Delta S_E = 0$, the monotonicity of f and the logarithm, as well as $I_{S:E} \leq S(\rho') - S(\rho)$. This shows that, for any fixed ΔV_S and fixed system dimension d_S, d_E has to grow as $d_E \geq O[\exp(\delta^{-1/8})]$ for the above equation to be satisfied. For state transitions where the entropy change is small, reducing the variance is therefore possible only at the expense of using a large catalyst.

F. Local monotonicity and entropy

Result 4 is nontrivial only when the rhs of Eq. (8) is positive (i.e., when there is significant decrease of marginal variances compared to the mutual information). This is because the lhs is always non-negative—a property we call *local monotonicity with respect to maximally mixed states (and unital channels)*. The local monotonicity of entropy follows straightforwardly from the fact that it is Schur concave, additive, and subadditive. Our last result is to show that, conversely, this property is essentially *unique* to the

von Neumann entropy, namely, it singles out the latter from all continuous functions on quantum states.

To state our result, let us first define local monotonicity more formally. Consider a function f on quantum states on finite-dimensional Hilbert spaces. Let $\mathbb{I}_1$ and $\mathbb{I}_2$ be maximally mixed states on systems S_1 and S_2, and let $\mathcal{U}$ be a unital channel, namely $\mathcal{U}(\mathbb{I}_1 \otimes \mathbb{I}_2) = \mathbb{I}_1 \otimes \mathbb{I}_2$. We say that f is *locally monotonic with respect to maximally mixed states* if for any such $\mathbb{I}_i$, $\mathcal{U}$ and two states $\rho_i \in \mathcal{D}(S_i)$, we have

$$f(\rho_1) + f(\rho_2) \leq f(\rho'_1) + f(\rho'_2),$$

where $\rho'_1 = \text{tr}_2 \mathcal{U}(\rho_1 \otimes \rho_2)$ and similarly for ρ'_2. We then have the following result.

Result 5. (Uniqueness of von Neumann entropy). *Let f be a continuous function that is locally monotonic with respect to maximally mixed states. Then*

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

where S is the von Neumann entropy, d is the Hilbert-space dimension of ρ and b_d depends only on d but not otherwise on ρ. It is sufficient for this result to restrict the set of channels to unitary channels.

The proof of the result can be found in Appendix H. While there exist many axiomatic characterizations of the entropy, we consider the above interesting because its only axiom (apart from continuity)—local monotonicity—is directly motivated by physical, as opposed to mathematical, considerations. This is because physics is often concerned with the possible changes of local quantities in the course of physical processes. As an example, in Box 4, we apply Result 5 to derive a version of Boltzmann's H theorem.

Finally, going back to Result 4, we see that just like how Result 3 provides a strengthening to the monotonicity of the entropy, Result 4 provides a strengthening to the local monotonicity of the entropy.

Box 4. A version of Boltzmann's H theorem

Here we note that one can derive a version of Boltzmann's H *theorem* as a simple corollary of Result 5: consider a "gas" of N independent quantum systems initially in state $\rho^{(0)} = \otimes_{i=1}^N \rho_i$. At any point in time t, two of these systems, call them i and j, first undergo a joint evolution, described by a (possibly random) unitary channel $\mathcal{U}_t$. We further assume that, following this interaction, any correlations between these two particles vanish (this is the infamous "Stoßzahlansatz"). A single iteration of

this process then yields the chain of states

$$\rho_{ij}^{(t)} = \rho_i \otimes \rho_j \rightarrow \rho_{ij}' = \mathcal{U}_t(\rho_{ij}^{(t)}) \rightarrow \rho_{ij}^{(t+1)} = \rho_i' \otimes \rho_j',$$

where $\rho_{ij} = \mathrm{tr}_{(ij)^c}[\rho]$. All other systems in the gas remain unchanged during this process. We are now interested in finding a continuous, real-valued function f such that

$$f(\rho^{(t)}) \leq f(\rho^{(t+1)})$$

for all t and all possible $\mathcal{U}_t$. Then, the above result implies that f exists and is given by the von Neumann entropy.

III. MAIN RESULTS FOR GENERIC QUANTUM CHANNELS

In the previous section, we have provided an overview of our main results for the special case of the resource theory of purity. We now turn to an exposition of our results in their full generality. All previously mentioned results are special cases of the general results presented here. Since we discussed the interpretation of the results already in the last section, we now focus on the technical and formal presentation. Some of technical proofs are nevertheless delegated to the appendices.

A. Notation and main concepts

Recall the preorder $\succ$ defined at the beginning of the last section for pairs of states (ρ, σ). A well-known example of this ordering that goes beyond majorization is β majorization in quantum thermodynamics, where $\sigma = \sigma'$ is the thermal state of the system at inverse temperature β [19]. Throughout the following, we focus on the *quasiclassical* setting, in which the initial and final pairs commute with one another, i.e., $[\rho, \sigma] = [\rho', \sigma'] = 0$. This significantly simplifies the setting mathematically, while still allowing some genuine quantum features, for example, when $\sigma = \sigma'$ but $[\rho, \rho'] \neq 0$.

1. Lorenz curves

A key tool in proving our sufficiency results is a well-known connection between the preorder $\succ$ and the *Lorenz curve*. For self-consistency, we present this connection and all relevant constructions in the notation of this paper. In particular, let ρ and σ be two commuting positive-semidefinite operators on the same d-dimensional Hilbert space $\mathcal{H}$, and denote by $\{|i\rangle\}_{i=1}^d$ an orthonormal basis of $\mathcal{H}$ that simultaneously diagonalizes both ρ and σ, i.e., $\sigma = \sum_i s_i |i\rangle\langle i|$ and $\rho = \sum_i p_i |i\rangle\langle i|$. Furthermore, let us assume without loss of generality that the basis $\{|i\rangle\}_{i=1}^d$ *orders ρ*

relative to σ, namely

$$\frac{p_i}{s_i} \geq \frac{p_{i+1}}{s_{i+1}} \tag{9}$$

for any $i = 1, \ldots, d$. Note that in this ordering neither the $(p_i)_i$ nor $(s_i)_i$ are necessarily ordered. Given the notations introduced, we can now introduce the *Lorenz curve*.

Definition 3: (Lorenz curves). Given two commuting quantum states ρ and σ, the Lorenz curve $\mathcal{L}_{\rho|\sigma}(x) : [0, 1] \rightarrow [0, 1]$ is given by the piecewise linear curve that connects the points

$$\left\{ \sum_{i=1}^k s_i, \sum_{i=1}^k p_i \right\}_{k=1}^d.$$

If ρ and σ do not commute, the Lorenz curve is taken to be $\mathcal{L}_{\rho|\sigma} = \mathcal{L}_{\mathcal{W}(\rho)|\sigma}$, where $\mathcal{W}(\rho)$ is the state pinched to the eigenbasis of σ, i.e., $\mathcal{W}(\rho) = \sum_i P_i \rho P_i$, with P_i the projectors onto the eigenspaces of σ.

Due to the way we have ordered the eigenvalues according to Eq. (9), the Lorenz curve is by definition always concave. The following now provides a simple and well-known equivalence relation between Lorenz curves and σ majorization.

Theorem 4. *Given two pairs of commuting states (ρ, σ) and (ρ', σ'), the following are equivalent:*

1. *For the entire range of $x \in [0, 1]$,*

$$\mathcal{L}_{\rho|\sigma}(x) \geq \mathcal{L}_{\rho'|\sigma'}(x).$$

2. *$(\rho, \sigma) \succ (\rho', \sigma')$.*

Theorem 4 is a condensed version of a more extended statement found as Theorem 24 in Appendix B, also see Refs. [13–16].

2. Flat and steep approximations relative to σ

We are now in a position to define the following approximations, known as flat and steep approximations of a state ρ relative to σ, denoted as $\rho_{\mathrm{fl}}^\epsilon$ and $\rho_{\mathrm{st}}^\epsilon$ respectively, which will play an important role for the derivation of our results. These states were initially defined in Ref. [24] for the special case of thermal reference states. Although the following Definitions 5 and 6 seem technical, they have the essential appealing property that for any state ρ and any $1 > \epsilon > 0$, we have [80]

$$\rho_{\mathrm{st}}^\epsilon \succ_\sigma \rho \succ_\sigma \rho_{\mathrm{fl}}^\epsilon.$$

The states are constructed as follows.

Definition 5: (Flat approximation relative to σ). Let σ, ρ be commuting quantum states on a d-dimensional Hilbert space $\mathcal{H}$ and $\{|i\rangle\}_{i=1}^d$ a common eigenbasis of the two states that orders ρ relative to σ, yielding $\sigma = \sum_i s_i |i\rangle\langle i|$ and $\rho = \sum_i p_i |i\rangle\langle i|$. For any $0 \le \epsilon \le 1$, the ϵ-*flattest approximation relative to* σ is the state $\rho_{\mathrm{fl}}^\epsilon = \sum_i \bar{p}_i |i\rangle\langle i|$, where the $\bar{p}_i$ are defined as follows: if $D(\rho, \sigma) < \epsilon$, set $\bar{p}_i = s_i$. Otherwise, define $M \in \{1, 2, \ldots, d-1\}$ as the smallest integer such that

$$\epsilon \le \sum_{i=1}^M p_i - \frac{p_{M+1}}{s_{M+1}} \sum_{i=1}^M s_i$$

and let $N \in \{2, \ldots, d\}$ be the largest integer such that

$$\epsilon \le \frac{p_{N-1}}{s_{N-1}} \sum_{i=N}^d s_i - \sum_{i=N}^d p_i.$$

These integers always exist when $\epsilon \le D(\rho, \sigma)$ and moreover satisfy $M \le N$ ([24], Appendix D, Lemma 6). Using these definitions, finally set

$$\bar{p}_i = \begin{cases} s_i \dfrac{(\sum_{j=1}^M p_j) - \epsilon}{\sum_{j=1}^M s_j}, & \text{if } i \le M \\[3mm] s_i \dfrac{(\sum_{j=N}^d p_j) + \epsilon}{\sum_{j=N}^d s_j}, & \text{if } i \ge N \\[3mm] p_i & \text{otherwise.} \end{cases}$$

Definition 6: (Steep approximation relative to σ). Let σ, ρ be commuting quantum states on a d-dimensional Hilbert space $\mathcal{H}$ and $\{|i\rangle\}_{i=1}^d$ a common eigenbasis of the two states that orders ρ relative to σ, yielding $\sigma = \sum_i s_i |i\rangle\langle i|$ and $\rho = \sum_i^d p_i |i\rangle\langle i|$. Then, for $0 \le \epsilon \le 1$, the ϵ-*steep approximation relative to* σ is the state $\rho_{\mathrm{st}}^\epsilon = \sum_i \hat{p}_i |i\rangle\langle i|$, such that if $\epsilon \le 1 - p_1$,

$$\hat{p}_i = \begin{cases} p_1 + \epsilon, & \text{if } i = 1 \\ p_i, & \text{if } 1 < i < R \\ p_i - (\epsilon - r), & \text{if } i = R \\ 0 & \text{otherwise,} \end{cases}$$

where $R \in \{2, \ldots, d\}$ is the largest index such that $\sum_{i=R}^d p_i \ge \epsilon$ and by definition of R, we have $r = \sum_{i=R+1}^d p_i \le \epsilon$. On the other hand, if $\epsilon > 1 - p_1$, define

$$\hat{p}_i = \begin{cases} 1, & \text{if } i = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Figure 2 illustrates an example of steep and flat approximations. We now state a key technical lemma, that provides the properties of Lorenz curves for ϵ-steep and ϵ-flat approximations for any pair (ρ, σ).
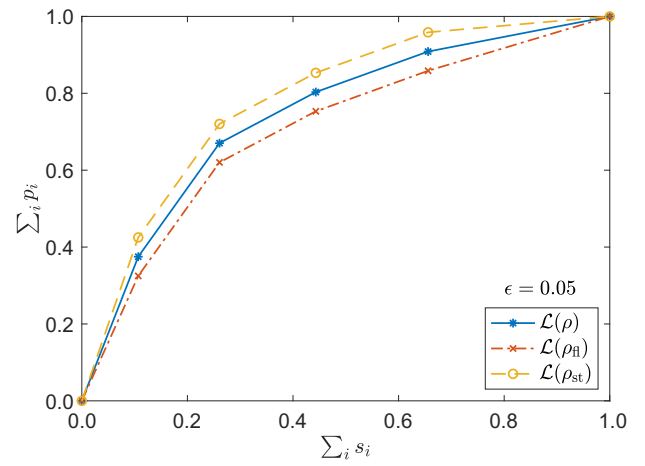

FIG. 2. An example illustrating the Lorenz curves $\mathcal{L}(\rho), \mathcal{L}(\rho_{\mathrm{st}}^\epsilon)$, and $\mathcal{L}(\rho_{\mathrm{fl}}^\epsilon)$, for some $\epsilon = 0.05$ and a state ρ of rank 5. The construction according to Definitions 5 and 6 are so that $\mathcal{L}(\rho_{\mathrm{st}}^\epsilon) \ge \mathcal{L}(\rho) \ge \mathcal{L}(\rho_{\mathrm{fl}}^\epsilon)$ for any value $\epsilon \in [0, 1]$ and any state ρ.

Lemma 7. *Let (ρ, σ) be two commuting quantum states, with σ full rank and let $1 > \epsilon > 0$. Then, for $x \in [0, 1]$,*

$$\mathcal{L}_{\rho_{\mathrm{st}}^\epsilon | \sigma}(x) \ge \ell_{r_{\mathrm{st}}}(x), \quad r_{\mathrm{st}} = 2^{S(\rho \| \sigma) - f_\sigma(\rho, \epsilon)},$$

$$\mathcal{L}_{\rho_{\mathrm{fl}}^\epsilon | \sigma}(x) \le \ell_{r_{\mathrm{fl}}}(x), \quad r_{\mathrm{fl}} = 2^{S(\rho \| \sigma) + f_\sigma(\rho, \epsilon)},$$

where $f_\sigma(\rho, \epsilon) := \sqrt{V(\rho \| \sigma)(\epsilon^{-1} - 1)}$ and $l_c(x) = \min(c \cdot x, 1)$.

This lemma is proven in Appendix E. Intuitively, it shows that the steep (flat) approximations allow us to obtain a state close to ρ in trace distance, with its Lorenz curve being lower (upper) bounded by straight lines l_c with gradients $c = r_{\mathrm{st}}(r_{\mathrm{fl}})$ governed by both the relative entropy and its variance. These simple bounds on the Lorenz curves of ρ_{st} and ρ_{fl} are crucial for our derivation of Theorems 8 and 9, which are the general statements for Results 1 and 2 and are obtained as a direct consequence of this Lemma 7.

B. Sufficient criteria for state transitions under σ majorization

Using Lemma 7, we can now derive sufficiency conditions for approximate state transitions between commuting pairs of quantum states.

Theorem 8. *Let (ρ, σ) and (ρ', σ') be two pairs of commuting states with σ, σ' both full rank. For $1 > \epsilon > 0$, let $f_\sigma(\rho, \epsilon) := \sqrt{V(\rho \| \sigma)(2\epsilon^{-1} - 1)}$. If*

$$S(\rho \| \sigma) - f_\sigma(\rho, \epsilon) \ge S(\rho' \| \sigma') + f_{\sigma'}(\rho', \epsilon),$$

then $(\rho, \sigma) \succ_\epsilon (\rho', \sigma')$.

Proof. Let $\bar{\epsilon} = \epsilon/2$, and $r_{\text{st}} = 2^{S(\rho\|\sigma)-f_\sigma(\rho,\bar{\epsilon})}$, and $r'_{\text{fl}} = 2^{S(\rho'\|\sigma')+f_{\sigma'}(\rho',\bar{\epsilon})}$. Note that if the above condition holds, then in the whole range of $x \in [0,1]$, we have that $l_{r_{\text{st}}} \geq l_{r'_{\text{fl}}}$. By Lemma 7, we then have that in that range

$$\mathcal{L}_{\rho_{\text{st}}^{\bar{\epsilon}}|\sigma}(x) \geq \ell_{r_{\text{st}}}(x) \geq l_{r'_{\text{fl}}}(x) \geq \mathcal{L}_{\rho_{\text{fl}}'^{\bar{\epsilon}}|\sigma'}(x),$$

which by Theorem 4 implies that there exists a channel $\mathcal{E}$ such that $\mathcal{E}(\sigma) = \sigma'$ and $\mathcal{E}(\rho_{\text{st}}^{\bar{\epsilon}}) = \rho_{\text{fl}}'^{\bar{\epsilon}}$. Applying the same channel to ρ yields a state $\mathcal{E}(\rho) = \hat{\rho}'$ such that $D(\hat{\rho}', \rho') \leq \epsilon$, since

$$D(\hat{\rho}', \rho') \leq D(\hat{\rho}', \rho_{\text{fl}}'^{\bar{\epsilon}}) + D(\rho_{\text{fl}}'^{\bar{\epsilon}}, \rho')$$
$$\leq D(\rho, \rho_{\text{st}}^{\bar{\epsilon}}) + \bar{\epsilon} \leq 2\bar{\epsilon} = \epsilon.$$

∎

Result 1 in Sec. II follows as a special case for $\sigma = \mathbb{I}$. As mentioned earlier, in Appendix F, we apply Theorem 8 to derive sufficient conditions for IID state transitions with large but finite number of states, recovering a previously observed resonance condition, where second-order corrections can vanish even for nonzero variances of the initial and final states ρ, ρ'.

C. Relation to smoothed min- and max-relative entropies

Two quantities that are useful in describing single-shot processes are the min- and max-relative entropy. Given a positive semidefinite operator $\sigma \geq 0$ and a quantum state $\rho \in \mathcal{D}(S)$, let π_ρ denote the projector onto the support of ρ. Moreover, for two operators A and B, we write $A \geq B$ to mean that the operator $A - B$ is positive semidefinite. In terms of this notation, if $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, then we have the following definitions [81]:

$$S_{\min}(\rho\|\sigma) := -\log\text{tr}(\pi_\rho\sigma),$$
$$S_{\max}(\rho\|\sigma) := \log\min\{\lambda : \rho \leq \lambda\sigma\}.$$

The *smoothed* variants are further defined as

$$S_{\min}^\epsilon(\rho\|\sigma) := \max_{\tilde{\rho}\in\mathcal{B}^\varepsilon(\rho)} S_{\min}(\tilde{\rho}||\sigma),$$
$$S_{\max}^\epsilon(\rho\|\sigma) := \min_{\tilde{\rho}\in\mathcal{B}^\varepsilon(\rho)} S_{\max}(\tilde{\rho}||\sigma),$$

where the optimizations are over the set of all quantum states ε close in terms of trace distance to ρ, denoted as $\mathcal{B}^\varepsilon(\rho)$. Finally, define as $S_{\max}^\epsilon(\rho) := \log(d) - S_{\min}^\epsilon(\rho\|\mathbb{I})$ and $S_{\min}^\epsilon(\rho) := \log(d) - S_{\max}^\epsilon(\rho\|\mathbb{I})$ the min and max entropies utilized in Result 2.

We now present the generalization of Result 2 for the smoothed min- and max-relative entropies, which is easily proven by making use of Lemma 7.

Theorem 9. *(Bounds on smoothed Rényi divergences). Given $\rho, \sigma \in \mathcal{D}(S)$, let $1 > \epsilon > 0$. Then,*

$$S_{\max}^\epsilon(\rho\|\sigma) - S(\rho\|\sigma) \leq f_\sigma(\rho, \epsilon),$$
$$S(\rho\|\sigma) - S_{\min}^\epsilon(\rho\|\sigma) \leq f_\sigma(\rho, \epsilon),$$

where $f_\sigma(\rho, \epsilon) := \sqrt{V(\rho\|\sigma)(\epsilon^{-1}-1)}$.

Proof. We know from Ref. [24] that $S_{\max}^\epsilon(\rho\|\sigma) = S_{\max}(\rho_{\text{fl}}^\epsilon\|\sigma)$. Therefore,

$$S_{\max}^\epsilon(\rho\|\sigma) - S(\rho\|\sigma) = S_{\max}(\rho_{\text{fl}}^\epsilon\|\sigma) - S(\rho\|\sigma) \leq f_\sigma(\rho, \epsilon).$$

The last inequality follows from Lemma 7, which implies

$$S_{\max}(\rho_{\text{fl}}^\epsilon\|\sigma) \leq \log r_{\text{fl}} = S(\rho\|\sigma) + f_\sigma(\rho, \epsilon),$$

since $S_{\max}(\rho\|\sigma)$ is simply the logarithm of the gradient of the Lorenz curve $\mathcal{L}_{\rho|\sigma}$ at the origin. On the other hand,

$$S_{\min}^\epsilon(\rho\|\sigma) = \max_{\tilde{\rho}\in\mathcal{B}^\varepsilon(\rho)} S_{\min}(\tilde{\rho}\|\sigma) \geq S_{\min}(\rho_{\text{st}}^\epsilon\|\sigma). \quad (10)$$

Now, let π_ρ denote the projector onto the support of ρ. By definition of the Lorenz curve we have that

$$\text{tr}(\pi_\rho\sigma) = \min\{x|x \in [0,1], \mathcal{L}_{\rho|\sigma}(x) = 1\}.$$

Using this fact and Lemma 7, in particular the definition of $l_c(x)$, we then find

$$\text{tr}(\pi_{\rho_{\text{st}}^\epsilon}\sigma) \leq r_{\text{st}}^{-1}.$$

Combining this with the definition of the smooth min-relative entropy then yields

$$S_{\min}^\epsilon(\rho\|\sigma) \geq S_{\min}(\rho_{\text{st}}^\epsilon\|\sigma) \geq \log r_{\text{st}} = S(\rho\|\sigma) - f_\sigma(\rho, \epsilon). \quad (11)$$

Finally, combining Eqs. (10) and (11) yields the second claim of the theorem. ∎

D. Uniform continuity and correction to subadditivity

We here show that the variance of relative surprisal is uniformly continuous and bound its violation of subadditivity. Both of these properties of the relative variance are key tools to derive our main results, however we believe that they are of independent interest and use.

Lemma 10. *(Uniform continuity of the relative variance). Let σ be a positive-definite operator with $0 < \sigma \leq \mathbb{I}$ and smallest eigenvalue $s_{\min}$ on a d-dimensional Hilbert space*

with $d \geq 2$, and ρ, ρ' be two states that both commute with σ. Then,

$$|V(\rho\|\sigma) - V(\rho'\|\sigma)| \leq 2K\sqrt{D(\rho,\rho')},$$

where $K = 8\log^2(d) + \log(d) + 2\log^2(s_{\min}) - 4\ln(2)$
$\log(s_{\min}) + 15$.

This lemma is proven in Appendix C.

Lemma 11. (Correction to subadditivity of relative variance.) *Let ρ, σ be two commuting quantum states on a d-dimensional, bipartite system, with $d \geq 2$ and σ full rank. If $\sigma = \sigma_1 \otimes \sigma_2$ is a product state with smallest eigenvalue $s_{\min}$, then*

$$V(\rho\|\sigma) \leq V(\rho_1\|\sigma_1) + V(\rho_2\|\sigma_2) + K'f(I_{1:2}),$$

where $K' = \sqrt{2\ln(2)}\left[12 + \log(s_{\min})^2 + 8\log^2(d)\right]$, $f(x) = \max\{\sqrt[4]{x}, \sqrt{x}\}$, and $I_{1:2}$ denotes the mutual information between the two partitions of ρ.

This lemma is proven in Appendix D.

E. A new monotone and relative entropy production

We now turn to the presentation and derivation of the results that generalize Results 3 and 4. We begin by noting that the relative entropy $S(\rho\|\sigma)$ is a nonincreasing resource monotone with respect to the ordering $\succ$, generalizing the Schur concavity of the von Neumann entropy. We then have the following generalization of Lemma 2, which we prove in Appendix G.

Theorem 12. *Let (ρ, σ) and (ρ', σ') be two pairs of commuting quantum states, with σ, σ' both full rank. If $(\rho, \sigma) \succ (\rho', \sigma')$, then it holds that $M_{s_{\min}}(\rho'\|\sigma') \geq M_{s_{\min}}(\rho\|\sigma)$, where*

$$M_x(\rho\|\sigma) := V(\rho\|\sigma) + \left(\frac{1}{\ln(2)} - \log(x) - S(\rho\|\sigma)\right)^2,$$

and $s_{\min}$ denotes the smallest eigenvalue of σ.

Result 2 follows by setting $\sigma = \sigma' = \mathbb{I}$. Furthermore, since σ is the minimum of the preorder $\succ_\sigma$, monotonicity implies that

$$0 \leq M_{s_{\min}}(\rho\|\sigma) \leq M_{s_{\min}}(\sigma\|\sigma) = \left[\frac{1}{\ln(2)} - \log(s_{\min})\right]^2$$

for any state ρ that commutes with σ.

We now derive the following corollary of the above theorem as a general version of Result 3, where we write $\Delta S = S(\rho\|\sigma) - S(\rho'\|\sigma')$ and $\Delta V = V(\rho\|\sigma) - V(\rho'\|\sigma')$.

Corollary 13. *Let (ρ, σ) and (ρ', σ') be pairs of commuting states, with σ, σ' both full rank. If $(\rho, \sigma) \succ (\rho', \sigma')$, then*

$$\Delta S \geq \frac{\Delta V}{2\sqrt{M_{s_{\min}}(\rho\|\sigma)}} \geq \frac{\Delta V}{2[1/\ln(2) - \log(s_{\min})]}.$$

Proof. By monotonicity of the relative entropy and positivity of M, the statement is trivially true whenever $\Delta V \leq 0$. Hence, assume that $\Delta V > 0$. By Theorem 12, we know that

$$M_{s_{\min}}(\rho'\|\sigma') \geq M_{s_{\min}}(\rho\|\sigma).$$

Let us write $a = 1/\ln(2) - \log(s_{\min})$, so that

$$M_{s_{\min}}(\rho\|\sigma) = V(\rho\|\sigma) + [a - S(\rho\|\sigma)]^2.$$

Then inserting the definition of $M_{s_{\min}}$ and reshuffling terms yields [using $x^2 - y^2 = (x-y)(x+y)$]

$$\begin{aligned}
0 &\leq \Delta S \cdot \left[2a - S(\rho\|\sigma) - S(\rho'\|\sigma')\right] - \Delta V \\
&= \Delta S \cdot \left[2a - 2S(\rho\|\sigma) + S(\rho\|\sigma) - S(\rho'\|\sigma')\right] - \Delta V \\
&= \Delta S \cdot \left[2a - 2S(\rho\|\sigma) + \Delta S\right] - \Delta V \\
&= (\Delta S)^2 + 2\chi(\Delta S) - \Delta V,
\end{aligned}$$

where we write $\chi = a - S(\rho\|\sigma) \geq 0$. Solving the quadratic equation in ΔS then gives

$$\begin{aligned}
\Delta S &\geq -\chi + \sqrt{\chi^2 + \Delta V} \\
&= -\sqrt{\chi^2} + \sqrt{\chi^2 + \Delta V} \\
&\geq \frac{\Delta V}{2\sqrt{\chi^2 + \Delta V}} \\
&\geq \frac{\Delta V}{2\sqrt{\chi^2 + V(\rho\|\sigma)}} = \frac{\Delta V}{2\sqrt{M_{s_{\min}}(\rho\|\sigma)}}.
\end{aligned}$$

Here, we use the fact that $\Delta S \geq 0$ by monotonicity in the first step (to disregard one solution), positivity of χ in the second step and the concavity of the square root in the third step [more precisely that $f(y) \geq f(x) + f'(y)(y - x)$ for any differentiable concave function]. This concludes the proof. ∎

We note in passing that such lower bounds on the production of relative entropy are essential for quantifying irreversibility in thermodynamics, where $(1/\beta)S(\rho\|\tau_\beta)$ denotes the nonequilibrium free energy of a system in state ρ in an environment of inverse temperature β. Here, τ_β denotes the Gibbs state of the system at inverse temperature β. We leave the detailed investigation of applications of our results to thermodynamics for future work.

Next, we present the generalized version of Result 4. Let S and E be two systems of respective dimension d_S

and d_E and $\sigma \equiv \sigma_S \otimes \sigma_E$ as well as $\sigma' \equiv \sigma'_S \otimes \sigma'_E$ be two fixed product states on the joint system SE. Let $\mathcal{E} : \mathcal{D}(S \otimes E) \to \mathcal{D}(S \otimes E)$ be a quantum channel such that $\mathcal{E}(\sigma) = \sigma'$. Using this channel we can define a channel $\mathcal{C} : \mathcal{D}(S) \to \mathcal{D}(S)$ as

$$\mathcal{C}(\cdot) = \mathrm{tr}_E[\mathcal{E}(\cdot \otimes \rho_E)], \qquad (12)$$

for an initial state ρ_E of the environment. As in the previous section, for some initial state ρ_S on S, we denote as $\rho'_S = \mathcal{C}(\rho_S)$ the final state on S, as $\Delta S_S = S(\rho_S \| \sigma_S) - S(\rho'_S \| \sigma'_S)$ the marginal change of relative entropy on S, and similarly for ΔV_S and the environment E. Finally, $I_{S:E}$ is the mutual information of the final state $\mathcal{E}(\rho_S \otimes \rho_E)$, as defined in Eq. (7) (with $\mathcal{U}$ replaced by $\mathcal{E}$). We then have the following.

Theorem 14. *Let $\mathcal{C}$ be a channel defined via Eq. (12), for fixed and full-rank states σ, σ' as well as ρ_E. Then, for any initial system state ρ_S such that $\rho_S \otimes \rho_E$ commutes with σ and $\mathcal{E}(\rho_S \otimes \rho_E)$ commutes with σ', we have*

$$\Delta S_S + \Delta S_E \geq \frac{\Delta V_S + \Delta V_E - K f(I_{S:E})}{2 \sqrt{M_{s_{\min}}(\rho_S \otimes \rho_E \| \sigma)}},$$

where

$$K = \sqrt{2 \ln(2)} \left[12 + \log^2(s_{\min}) + 8 \log^2(d_S d_E) \right]$$

and $f(x) = \max\{\sqrt[4]{x}, x^2\}$. Here, $s_{\min}$ is the smallest eigenvalue of σ.

Proof. Applying Corollary 13 with $\rho \equiv \rho_S \otimes \rho_E$ and Lemma 11 yields

$$S(\rho_S \otimes \rho_E \| \sigma) - S[\mathcal{C}(\rho_S \otimes \rho_E \| \sigma)]$$
$$\geq \frac{\Delta V_S + \Delta V_E - c f(I_{S:E})}{2 \sqrt{M(\rho_S \otimes \rho_E \| \sigma)}}.$$

The statement then follows from the fact that, for any state ρ'_{SE} on SE with mutual information $I_{S:E}$,

$$S(\rho'_{SE} \| \sigma') = S(\rho'_S \| \sigma'_S) + S(\rho'_E \| \sigma'_E) + I_{S:E}$$
$$\geq S(\rho'_S \| \sigma'_S) + S(\rho'_E \| \sigma'_E).$$

∎

F. Relative entropy from local monotonicity

Lastly, let us discuss the general version of local monotonicity, which uniquely characterizes the relative entropy. To do this, let $\mathcal{F}$ be the set of all finite-dimensional density matrices with full rank and let $\mathcal{C}_\mathcal{F}$ be the set of quantum channels that map states of full rank to states of full rank (on possibly different Hilbert spaces), symbolically $\mathcal{C}_\mathcal{F}(\mathcal{F}) \subseteq \mathcal{F}$. We further generalize the notion of local monotonicity to states $\sigma_1 \otimes \sigma_2$ that are not fixed points of a given channel: We say that a function f on pairs of quantum states (ρ, σ), with ρ defined on the same Hilbert space as $\sigma \in \mathcal{F}$, is *locally monotonic with respect to $\mathcal{C}_\mathcal{F}$* if $C[\sigma_1 \otimes \sigma_2] = \sigma'_1 \otimes \sigma'_2 \in \mathcal{F}$ for $C \in \mathcal{C}_\mathcal{F}$ and $\sigma_1 \otimes \sigma_2 \in \mathcal{F}$ implies

$$f(\rho_1, \sigma_1) + f(\rho_2, \sigma_2) \geq f(\rho'_1, \sigma'_1) + f(\rho'_2, \sigma'_2),$$

where again $\rho'_1 = \mathrm{tr}_2[C(\rho_1 \otimes \rho_2)]$ and similarly for ρ'_2. We then have the following theorem.

Theorem 15. *Let f be a function that is locally monotonic with respect to $\mathcal{C}_\mathcal{F}$ and assume that $\rho \mapsto f(\rho, \sigma)$ is continuous for fixed $\sigma \in \mathcal{F}$. Then*

$$f(\rho, \sigma) = a S(\rho \| \sigma) + b,$$

where a and b are constants.

The proof can be found in Appendix H.

IV. CONCLUSIONS AND OUTLOOK

In this work we comprehensively studied formal properties of the variance of (relative) surprisal together with their applications to single-shot (quantum) information theory. Before closing, let us comment on the high-level motivation for this work and open avenues for further research. To do this, we restrict again to the case of unital channels ($\sigma = \mathbb{I}$) for simplicity.

As discussed throughout the paper, the von Neumann entropy quantifies information-theoretic tasks in the asymptotic limit. Conversely, the min and max entropies typically appear in the fully single-shot regime. All these quantities are special cases of the Rényi entropies

$$S_\alpha(\rho) = \frac{1}{1-\alpha} \log(\mathrm{tr}[\rho^\alpha]), \qquad \alpha \in (0,1) \cup (1,\infty),$$

with $S(\rho) = S_1(\rho) := \lim_{\alpha \to 1} S_\alpha(\rho)$, $S_{\min}(\rho) := \lim_{\alpha \to \infty} S_\alpha(\rho)$ and $S_{\max}(\rho) := \lim_{\alpha \to 0} S_\alpha(\rho)$. Indeed, we can consider the min and max entropies to be the end points of the "Rényi curve" $\alpha \mapsto S_\alpha(\rho)$. This curve encodes the full spectrum of a state (see below). Hence, roughly speaking, one can say that in the single-shot regime the full shape of the curve matters, while in the asymptotic IID limit only the point $\alpha = 1$ matters. From this point of view, the approach to single-shot information using min and max entropies rests on the observation that the end points of the (smoothed) curve capture many of the operationally relevant single-shot effects for a given state.

In contrast, the approach presented here quantifies single-shot effects by studying not the end points but rather

the neighborhood of the Rényi curve around $\alpha = 1$. To see this, consider the Taylor expansion of $S_\alpha(\rho)$ around $\alpha = 1 + x$. Then first performing the expansion and finally taking the limit $x \to 0$ yields

$$S_\alpha(\rho) = \sum_{n=1}^{\infty} \frac{\kappa^{(n)}}{n!} (1 - \alpha)^{n-1}, \qquad (13)$$

where $\kappa^{(n)}$ is the nth *cumulant of surprisal*. In Appendix I, we present the definition of the cumulants of surprisal as well as the derivation of Eq. (13) (also see Ref. [38]).

Equation (13) is interesting for a number of reasons. To begin with, we have $\kappa^{(1)} = S(\rho)$ and $\kappa^{(2)} = V(\rho)$. Hence, Eq. (13) shows that the variance of surprisal is (up to a factor of -2) the slope of the Rényi curve at $\alpha = 1$ and gives the first-order correction to the approximation $S_\alpha(\rho) \approx S(\rho)$. This fact is well known [38,82]. It lets us apply some of our results to the Rényi curve. For instance, Result 2 relates the neighborhood of the Rényi curve around $\alpha = 1$ to its smoothed end points, while Result 3 constraints the possible changes of the Reńyi curve under unital channels, in the sense that the slope at $\alpha = 1$ can, by means of such channels, only be flattened at the expense of raising the curve at this point.

More generally, the expansion (13) is interesting because it implies that the higher-order cumulants of surprisal give a hierarchy of increasingly fine-grained knowledge about a state's spectrum. This follows once we recognize that for a d-dimensional state ρ, it suffices to know $S_n(\rho)$ for $n = 2, \ldots, d$ to fully reconstruct the spectrum of ρ (for the reader's convenience we provide a proof of this statement in Appendix J). In turn, the results of this paper, in which we studied the first order of this hierarchy, then suggest that studying the single-shot properties of higher-order cumulants or surprisal could yield insights about single-shot information theory that are somewhat complementary to the approach of smoothed Rényi entropies.

In particular, it would be interesting whether it is possible to construct a hierarchy of Schur-concave functions with increasing relevance at the single-shot level from cumulants of surprisal. As a first step in this direction, the following interesting problem arises: we have mentioned that knowing the Rényi entropies $S_n(\rho)$ for $n = 2, \ldots, d$ provides full information about the spectrum of the state. Is it also true that the first $d - 1$ cumulants of surprisal encode the full spectrum of the state? Another problem to consider is the extension of our study to the fully quantum setting of noncommuting matrices. We leave these questions to future work.

ACKNOWLEDGMENTS

APPENDIX A: OVERVIEW OF APPENDIX

Appendix B establishes all the notation used throughout our proofs and collects a handful of technical lemmas.

Appendices C and D present the proofs for Lemma 10 and Lemma 11, respectively.

Appendix E presents the proof of the central technical Lemma 7 that underlies all of our sufficiency results.

Appendix F presents the application of Theorem 8 to the case of finite IID sequences.

Appendix G then provides the proof of Theorem 12.

Appendix H discusses details on the axiomatic characterization of locally monotonic functions, including the proofs of Result 5 and Theorem 15.

Finally, Appendix I provides the details to the expansion Eq. (13) and Appendix J sketches the proof that a state's spectrum can be inferred from the values of $d - 1$ Rényi entropies, as claimed in the conclusion.

APPENDIX B: NOTATION AND AUXILIARY LEMMATA

In the following we make frequent use of the following definitions:

(a) $L(\rho \| \sigma) := \text{tr}\{\rho[\log(\rho) - \log(\sigma)]^2\}$;
(b) $\chi(x \| q) := x \log^2(x/q)$, defined for $q > 0$, and over the regime $x \in [0, 1]$ by continuous extension;
(c) $\eta(x) := -x \log(x)$, defined over the interval $[0, 1]$ by continuous extension;
(d) $h_b(x) := \eta(x) + \eta(1 - x)$, the binary entropy;
(e) $[d] := \{1, \ldots, d\}$.

We also remind the reader that we use logarithms with base 2, $\log = \log_2$. Lemmas 16–22 are technical tools used in the derivation of our results. We list them here for completeness.

Lemma 16. *For any $x \in [0, 1]$, $h_b(x) \leq 2 \ln(2) \sqrt{x(1 - x)}$.*

Lemma 17. (Klein's inequality). *Let ρ, σ be density operators. Then $S(\rho\|\sigma) \geq 0$ with equality if and only if $\rho = \sigma$.*

We also use the following generalization of the Fannes-Audenaert inequality, which is implied by Lemma 7 in Ref. [83].

Lemma 18. [Continuity of relative entropy (Lemma 7, Ref. [83])]. *Consider any full-rank state σ with $s_{\min} > 0$ denoting its smallest eigenvalue. Then, for any two states ρ, ρ' such that $D(\rho, \rho') \leq \epsilon$, we have*

$$|S(\rho\|\sigma) - S(\rho'\|\sigma)| \leq -\log(s_{\min})\epsilon + (1 + \epsilon)h_b\left(\frac{\epsilon}{1+\epsilon}\right).$$

Lemma 19. (Pinsker inequality). *For quantum states ρ, σ acting on the same Hilbert space, $S(\rho\|\sigma) \geq 1/2\ln(2) \|\rho - \sigma\|_1^2$*

Lemma 20. *Let $\mathcal{E}$ be a quantum channel such that $\mathcal{E}(\sigma) = \sigma'$, for two states $\sigma = \sum_i^d s_i|i\rangle\langle i|$ and $\sigma' = \sum_i^{d'} s_i'|i\rangle\langle i|$ (note that $d \neq d'$ in general). Furthermore, given a state $\rho = \sum_i^d p_i|i\rangle\langle i|$, suppose that $\rho' = \mathcal{E}(\rho) = \sum_i^{d'} q_i|i\rangle\langle i|$ commutes with σ'. Then there exists a right stochastic $d \times d'$ matrix E, that is, a matrix with all non-negative entries each of whose rows sums up to 1, such that*

$$pE = q,$$
$$sE = s',$$

where p, q, s, s' are simply vectors containing the eigenvalues of ρ, ρ', σ, and σ', respectively.

Lemma 21. (Cantelli-Chebyshev inequality). *Given a random variable X with finite mean μ, variance $\sigma^2 < \infty$ and $\lambda > 0$,*

$$\Pr(X - \mu \geq \lambda) \leq \frac{\sigma^2}{\sigma^2 + \lambda^2}.$$

Lemma 22. (Domination for definite integrals). *Let f, g be continuous functions. If $f(x) \geq g(x)$ in the interval $[a, b]$, then*

$$\int_a^b f(x)dx \geq \int_a^b g(x)dx.$$

Next, four technical lemmas are proven explicitly and used in later parts of our work.

Lemma 23. *Let ρ, ρ', and σ be three quantum states such that ρ and ρ' both commute with σ but not necessarily with another. Then there exists a unitary channel $\mathcal{U}$ such that (i) $[\rho, \mathcal{U}(\rho')] = 0$, (ii) $D[\rho, \mathcal{U}(\rho')] \leq D(\rho, \rho')$, and (iii) $\mathcal{U}(\rho')$ also commutes with σ.*

Proof. We write $\sigma = \oplus_i s_i \mathbb{1}_i$, where $\mathbb{1}_i$ is the identity operator in the ith eigenspace of σ. Similarly, we can write $\rho = \oplus_i \rho_i$ and $\rho' = \oplus_i \rho_i'$. We then have

$$D(\rho, \sigma) = \sum_i D(\rho_i, s_i\mathbb{1}_i),\ D(\rho', \sigma)$$
$$= \sum_i D(\rho_i', s_i\mathbb{1}_i),\ D(\rho, \rho') = \sum_i D(\rho_i, \rho_i').$$

The mapping $\rho' \mapsto U\rho'U^\dagger =: \mathcal{U}(\rho')$, with $U = \oplus_i U_i$ a block-diagonal unitary, is a σ-preserving quantum channel. Now, without loss of generality, choose a basis $|i, j\rangle$ in each eigenspace of σ such that $\rho_i = \sum_j p_{i,j}|i,j\rangle\langle i,j|$ with $p_{i,j} \geq p_{i,j+1}$. We can then choose U_i so that $U_i\rho_i'U_i^\dagger = \sum_i p_{i,j}'|i,j\rangle\langle i,j|$ with $p_{i,j}' \geq p_{i,j+1}'$ being the ordered eigenvalues of ρ_i'. Then clearly $[\rho_i, U_i\rho_i'U_i^\dagger] = 0$. Furthermore, collecting the respective eigenvalues in vectors $\mathbf{p}_i, \mathbf{p}_i'$, from Theorem 4 in Ref. [84] we directly find

$$D[\rho, \mathcal{U}(\rho')] = \sum_i D(\rho_i U_i \rho_i' U_i^\dagger) = \sum_i D(\mathbf{p}_i, \mathbf{p}_i')$$
$$\leq \sum_i D(\rho_i, \rho_i') = D(\rho, \rho').$$

∎

Theorem 24. (Relative-majorization and Lorenz curves). *Let $(p, q) \in \mathbb{R}_+^n$ and $(p', q') \in \mathbb{R}_+^m$ be two pairs of probability vectors (non-negative, normalized) such that when $q_i = 0$, then $p_i = 0$ and similarly for p' and q'. Furthermore, denote the states ρ, σ to be diagonalized in the same basis with eigenvalues p, q; while ρ', σ' are states diagonalized in the same basis with eigenvalues p', q', respectively. Then the following are equivalent:*

1. *There exists a $n \times m$-right stochastic matrix M such that $pM = p'$ and $qM = q'$.*
2. *$\mathcal{L}_{\rho|\sigma} \geq \mathcal{L}_{\rho'|\sigma'}$.*
3. *For every continuous concave function $g : \mathbb{R} \to \mathbb{R}$,*

$$\sum_{i=1}^n q_i' g\left(\frac{p_i'}{q_i'}\right) \geq \sum_{i=1}^m q_i g\left(\frac{p_i}{q_i}\right).$$

Proof. Forms and proofs of this theorem appear in various places of the literature, and can be found in, e.g., Refs. [13–16]. But since the proven statements often vary in some detail, we here present the proof for the reader's convenience.

$1. \Rightarrow 3.$: Let m_{ij} denote the elements of the matrix M that exists by assumption. Define the matrix A with elements $a_{ij} = q_i m_{ij}(q_j')^{-1}$. It is easy to check that A is a left-stochastic matrix and $Aq' = q$ and that $(p/q)A =$

(p'/q'), where p/q denotes elementwise division. Since p/q and p'/q' are real vectors due to fact that $q = 0$ implies $p = 0$, and similarly for the final pair, we can now apply Proposition A.1. p. 579 from Ref. [16] to arrive at the desired statement.

3. $\Rightarrow$ 2. : First, we note that, by definition of the Lorenz curve, $\mathcal{L}_{\rho|\sigma}(x) = \mathcal{L}_{\rho'|\sigma'}(x)$ for $x = 0, 1$. We hence need to show domination of the Lorenz curve only for $x \in (0, 1)$. For this, let $(g_t)_{t \in \mathbb{R}}$ and $(h_t)_{t \in \mathbb{R}}$ be families of parametrized functions defined by $g_t(x) := \max\{x - t, 0\}$ and $h_t(x) := \max\{t - x, 0\}$. By convexity of the max function, g_t as well as h_t are convex for every t. Next, define the function $S_{\rho|\sigma}$: $(0, 1) \rightarrow \mathbb{R}$ as

$$S_{\rho|\sigma}(x) := p_i/q_i, \quad \sum_{j=0}^{i-1} q_j < x \le \sum_{j=0}^{i} q_j,$$

where $q_0 = 0$ and we enumerate the vectors such that $p_i/q_i \ge p_{i+1}/q_{i+1}$, without loss of generality. A moment's thought will show that $S_{\rho|\sigma}(x)$ is the value of the slope of the Lorenz curve $\mathcal{L}_{\rho|\sigma}(x)$ at x, whenever this slope is well defined. Now, for fixed x, we distinguish between the case $S_{\rho|\sigma}(x) \ge S_{\rho'|\sigma'}(x)$ and $S_{\rho|\sigma}(x) < S_{\rho'|\sigma'}(x)$. In the first case, we evaluate condition 3. for g_t with $t = S_{\rho|\sigma}(x)$. This yields

$$\mathcal{L}_{\rho|\sigma}\left(\sum_{i=0}^{k} q_i\right) - t\sum_{i=0}^{k} q_i \ge \mathcal{L}_{\rho|\sigma}\left(\sum_{i=0}^{k'} q_i'\right) - t\sum_{i=0}^{k'} q_i',$$

where k is the largest index such that $p_k/q_k > t$, and similarly k' is the largest index such that $p_{k'}'/q_{k'}' > t$. Geometrically, we can interpret the above inequality as making a statement about two parallel lines with slope t running through the respective points $\mathcal{L}_{\rho|\sigma}(\sum_{i=1}^{k} q_i)$ (on the lhs) and $\mathcal{L}_{\rho|\sigma}(\sum_{i=1}^{k'} q_i)$ (on the rhs). Now, by our choice of t, the point $[x, \mathcal{L}_{\rho|\sigma}(x)]$ lies on the left-hand parallel, while by the fact that $S_{\rho|\sigma}(x) \ge S_{\rho'|\sigma'}(x)$, the point $(x, \mathcal{L}_{\rho'|\sigma'})$ is guaranteed to lie below the right-hand parallel. Hence, we have $\mathcal{L}_{\rho|\sigma}(x) \ge \mathcal{L}_{\rho'|\sigma'}(x)$, as required.

We now turn to the second case: if $S_{\rho|\sigma}(x) < S_{\rho'|\sigma'}(x)$, then we instead evaluate condition 3. for h_t with $t = S_{\rho'|\sigma'}(x)$. This yields, after canceling out some terms,

$$\mathcal{L}_{\rho|\sigma}\left(\sum_{i=0}^{k} q_i\right) - t\sum_{i=0}^{k} q_i \ge \mathcal{L}_{\rho|\sigma}\left(\sum_{i=0}^{k'} q_i'\right) - t\sum_{i=0}^{k'} q_i',$$

where in turn this time k is the largest index such that $p_k/q_k \ge t$, and similarly k' is the largest index such that $p_{k'}'/q_{k'}' \ge t$. By following the same argument as in the first case, we know that the point $[x, \mathcal{L}_{\rho'|\sigma'}(x)]$ lies on the right-hand parallel, while the point $(x, \mathcal{L}_{\rho'|\sigma'})$ is guaranteed to lie above the left-hand parallel. Hence, we have $\mathcal{L}_{\rho|\sigma}(x) \ge \mathcal{L}_{\rho'|\sigma'}(x)$, as required.
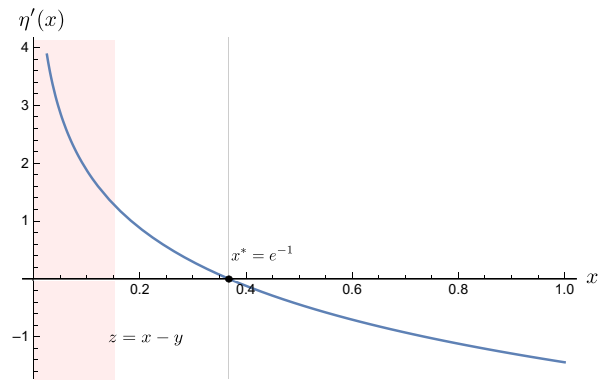

FIG. 3. The function $\eta'(x)$ and the corresponding area for the integral $F_z(0)$.

2. $\Rightarrow$ 1. : See Eq. 16 of Ref. [20] for the statement, where Ref. [14] contains the proof. ∎

Note that under the condition $[\rho, \sigma] = [\rho', \sigma'] = 0$, the existence of a quantum channel $\mathcal{E}$ such that $\mathcal{E}(\rho) \in \mathcal{B}^\epsilon(\rho')$, is equivalent to having a classical stochastic matrix M such that $pM = p'$ and $qM = q'$ (by Lemma 20). Theorem 4 is therefore implied by the first and second statements of Theorem 24.

Lemma 25. *Let* $x, y \in [0, 1]$ *such that* $|x - y| \le \frac{1}{2}$. *Then*

$$|\eta(x) - \eta(y)| \le \eta(|x - y|).$$

Proof. The statement is clearly true if $x = y$. Hence, without loss of generality let $x > y$, and set $z = |x - y| = x - y \le 1/2$. First, note that

$$|\eta(x) - \eta(y)| = \left| \int_0^z \eta'(y + r)dr \right| =: F_z(y),$$

and that $F_z(0) = \eta(z)$, so that is sufficient to show that,

$$F_z(0) \ge F_z(y). \tag{B1}$$

To show this, let us begin by evaluating the derivative $\eta'(x) = -1/\ln(2)[\ln(x) + 1]$, and noting that this is a monotonically decreasing function, with a root at $x^* = e^{-1}$. As graphically shown in Fig. 3, Eq. (B1) states that of all integrals of fixed width z, the one with the largest absolute value is the one over the interval $[0, z]$.

We now show that this is indeed the case. We consider three different cases: if $y \le x^* - z$, the statement is automatically true because η' is positive and monotonically decreasing below x^*, so that an application of Lemma 22

yields

$$F_z(0) = \int_0^z \eta'(r)dr \geq \int_0^z \eta'(y+r)dr = F_z(y).$$

From similar reasoning, we know that $F_z(y) \leq F_z(1-z)$ whenever $x^* \leq y \leq 1-z$, by monotonicity and negativity of η' above x^*. For the third case, when $x^* - z < y < x^*$, the fact that part of the integral is positive and part negative implies that

$$F_z(y) \leq \max\{F_z(x^*-z), F_z(x^*)\}$$
$$\leq \max\{F_z(0), F_z(1-z)\},$$

where in the second step we apply the bounds derived for the previous cases. Hence, it remains to show that $F_z(0) \geq F_z(1-z)$. To do so, first note that $F_z(0) = \eta(z)$ and $F_z(1-z) = \eta(1-z)$. Furthermore, the function $g(z) := \eta(z) - \eta(1-z)$ is continuous over $[0,1]$, is positive at $z = e^{-1} \in [0, 1/2]$, with roots at $x = 0, 1/2, 1$. By invoking the intermediate value theorem, we know that $g(z) \geq 0$ for all $z \leq 1/2$, which concludes the proof. ∎

Lemma 26. *Let $q \in (0,1]$ and $x, y, \in [0, q]$. If $|x - y| \leq q/e^2$, then*

$$|\chi(x\|q) - \chi(y\|q)| \leq \chi(|x - y| \| q).$$

Proof. The proof is in spirit very similar to that of Lemma 25. Again the statement is trivially true if $x = y$. Assume, then, again without loss of generality that $x > y$ and set $z = |x - y| = x - y \leq q/e^2$. We note that

$$|\chi(x\|q) - \chi(y\|q)| = \left| \int_0^z \chi'(y+r\|q)dr \right| =: G_z(y\|q),$$

and $G(0\|q) = \chi(z\|q)$, so it is sufficient to show $G_z(0\|q) \geq G_z(y\|q)$. Now, with the same strategy, let us first evaluate

$$\chi'(x\|q) = \frac{1}{\ln(2)^2} \left[2\ln(x/q) + \ln^2(x/q) \right],$$

and plot it in Fig. 4. We are interested in three intervals of this function: $[0, q/e^2]$, where it is monotonically decreasing and positive on the interval; $[q/e^2, q/e]$, where it is monotonically decreasing and negative; and $[q/e, q]$, where it is monotonically increasing and negative. By monotonicity on these separate intervals, the fact that $z \leq q/e^2$ (which implies that each of these intervals is at least
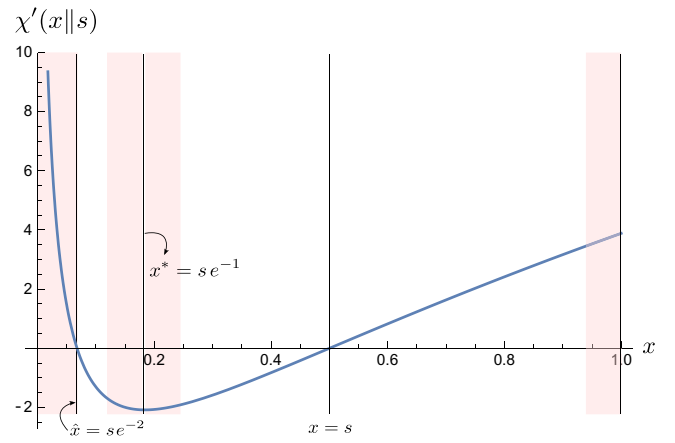

$\chi'(x\|s)$

FIG. 4. The function $\chi'(x\|q)$ for $q = 0.5$, with the relevant integrals highlighted in red.

as wide as z) and invoking Lemma 22, it follows that

$$\max_{r \in [0, s-z]} G_z(r\|q) = \max\{G_z(0\|q), G_z(q/e - z\|q)$$
$$+ G_z(q/e\|q)\},$$

by the following reasoning: for $r \leq q/e^2 - z$, by applying Lemma 22, we have that $G_z(r\|q) \leq G_z(0\|q)$ by positivity in that interval. Similarly, we can bound $G_z(r\|q)$ for the values $q/e^2 \leq r \leq s - z$ by the second term in the above bracket. Finally, for $q/e^2 - z < r < q/e^2$, parts of the integral cancel out, so that we can bound the integral by the above two terms. It hence remains to show that $G_z(0\|q)$ always dominates the second term above, which we check by explicit evaluation: we first have

$$G_z(0\|q) = \chi(z\|q) = \frac{z}{\ln^2(2)} \ln^2(z/q),$$

where, since $z \leq q/e^2 \leq 1$ by assumption and the fact that $\ln^2(x)$ is strictly monotonically decreasing in the range $[0,1]$, $\ln^2(z/q) \geq \ln^2(e^{-2}) = 4$. On the other hand, we can upper bound the second term by noting that

$$G_z(q/e - z\|q) + G_z(q/e\|q) \leq 2z\,|\chi'(q/e\|q)| = \frac{2z}{\ln^2(2)},$$

which is always smaller than $G_z(0\|q)$. ∎

APPENDIX C: PROOF OF UNIFORM CONTINUITY OF THE RELATIVE VARIANCE (LEMMA 10)

The main result of uniform continuity of relative variance (and therefore the nonrelative variance of surprisal) is proven in Lemma 10. To do so, let us first establish the following technical lemma.

Lemma 27. *Let σ be a positive-definite operator with $0 < \sigma \leq \mathbb{I}$ and smallest eigenvalue $s_{\min}$ on a d-dimensional Hilbert space with $d \geq 2$, and ρ, ρ' be two states that both commute with σ. If $D \equiv D(\rho, \rho') \leq 1/(2e^2)$, then*

$$|L(\rho\|\sigma) - L(\rho'\|\sigma)| \leq c_1 D \log^2(d) + \chi(2D\|1)$$
$$+ 2\eta(2D) \log(d),$$

where $c_1 = 12 + \log^2(s_{\min}) + 8\log^2(d)$. More generally, for any D,

$$|L(\rho\|\sigma) - L(\rho'\|\sigma)| \leq c_1 D \log^2(d) + c_2\sqrt{D},$$

with $c_2 = 6 + 2\log(d)$.

Proof. As the first step, we note that due to fact that both ρ and ρ' commute with σ, we need only to consider the spectra of the various states. This follows from Lemma 23. Let $\mathcal{U}$ denote the unitary channel defined in the proof of that Lemma. Then by construction we have that $[\rho, \mathcal{U}(\rho')] = 0$. Since

$$D[\rho, \mathcal{U}(\rho')] \leq D(\rho, \rho')$$

and $L[\mathcal{U}(\rho')\|\sigma] = L(\rho'\|\sigma)$, it follows that we can in the following replace ρ' by $\mathcal{U}(\rho')$, without loss of generality.

Since all three states then commute with another, we can assume the decompositions $\rho = \sum_i p_i |i\rangle\langle i|, \rho' = \sum_i q_i |i\rangle\langle i|$ and $\sigma = \sum_i s_i |i\rangle\langle i|$, in terms of which we have

$$D(\rho, \rho') = \frac{1}{2}\sum_i |p_i - q_i| = \frac{1}{2}\sum_i x_i,$$

$$|L(\rho\|\sigma) - L(\rho'\|\sigma)| = \left|\sum_i \chi(p_i\|s_i) - \chi(q_i\|s_i)\right|$$

$$\leq \sum_i |\chi(p_i\|s_i) - \chi(q_i\|s_i)|,$$

(C1)

and where we introduce the variable $x_i := |p_i - q_i|$. We now show that each of the terms in the rhs of Eq. (C1) can be upper bounded by a term of the form either $\chi(x_i\|s_i)$ or $C \cdot x_i$ for some constant C. To see this, consider the ith term in the sum and let us distinguish the following cases, where we assume without loss of generality that $q_i < p_i$.

Case I: $p_i \leq s_i/e^2$ In this case, we know that $x_i \leq s_i/e^2$ and so can apply Lemma 26 to find that

$$|\chi(p_i\|s_i) - \chi(q_i\|s_i)| \leq \chi(x_i\|s_i).$$

Case II: $s_i/e^2 \leq q_i$. In this case, we can make use of the fact that $\chi(\cdot\|s_i)$ is Lipschitz continuous in its first

argument over the interval $[s_i/e^2, 1]$. In particular, by differentiability of χ over this interval, we have that

$$|\chi(p_i\|s_i) - \chi(q_i\|s_i)| \leq \sup_{r\in[s_i/e^2,1]} |\chi'(r\|s_i)| \cdot |p_i - q_i|$$

$$= \max\{|\chi'(s_i/e\|s_i)|, |\chi'(1\|s_i)|\} \cdot x_i$$

$$= \frac{1}{\ln(2)^2} \max\{1, \ln^2(s_i) - 2\ln(s_i)\}$$

$$\cdot x_i =: \tilde{C}(s_i)x_i,$$

where $\tilde{C}(s_i)$ is the Lipschitz constant.

Case III: $q_i < s_i/e^2 < p_i$ Here, we distinguish three subcases. To discuss these cases, we note that since for fixed s_i, $\chi(\cdot\|s_i)$ is continuous and has roots at 0 and s_i, as well as a local maximum at s_i/e^2, by the mean value theorem there must be a point $q_i^* \in [s_i/e^2, s_i]$ such that

$$\chi(q_i\|s_i) = \chi(q_i^*\|s_i).$$

We now distinguish the following subcases: first, assume that $q_i^* \leq p_i$. Then we can make use of Lipschitz continuity, since

$$|\chi(p_i\|s_i) - \chi(q_i\|s_i)| = |\chi(p_i\|s_i) - \chi(q_i^*\|s_i)|$$

$$\leq \tilde{C}(s_i)|p_i - q_i^*| \leq \tilde{C}(s_i)(p_i - q_i^*)$$

$$\leq \tilde{C}(s_i)(p_i - q_i) = \tilde{C}(s_i)x_i.$$

Note that this subcase always covers situations in which $p_i \geq s_i$, because we are guaranteed that $q_i^* \leq s_i$. Hence, it remains to consider the case $q_i^* < p_i < s_i$. Now, if $x_i \leq s_i/e^2$, then we can apply Lemma 26 to find that

$$|\chi(p_i\|s_i) - \chi(q_i\|s_i)| \leq \chi(x_i\|s_i).$$

Finally, if $x_i > s_i/e^2$, then we have that

$$|\chi(p_i\|s_i) - \chi(q_i\|s_i)| = \chi(p_i\|s_i) - \chi(q_i\|s_i) < \chi(s_i/e^2\|s_i)$$

$$= 4s_i/e^2 < 4x_i,$$

where we twice use the fact that χ is strictly monotonically decreasing on the interval $[s_i/e^2, s_i]$ and positive. In the first step, combined with the definition of q_i^* this implies that $\chi(p_i\|s_i) > \chi(q_i\|s_i)$. In the second step, this implies that $\chi(p_i\|s_i) < \chi(s_i/e^2\|s_i)$.

Overall, we have seen that we can upper bound each of the terms on the rhs of Eq. (C1) by either $\chi(x_i\|s_i)$ or by $C(s_i)x_i$, where $\hat{C}(s_i) := \max\{4, \tilde{C}(s_i)\} = \max\{4, \chi'(1\|s_i)\}$.

In fact, we can derive a further, simple upper bound for $\hat{C}(s_i)$ as follows: first we have that $\chi'(1\|q) \geq 4$ for values $q \leq \exp[1 - \sqrt{1 + 4\ln^2(2)}] \approx 0.49$. At the same time, for values $q \leq \exp(-2/3) \approx 0.51$, we have that $\ln^2(q) -$

$2\ln(q) \le 4\ln^2(q)$. This implies that we have the upper bound

$$\hat{C}(s_i) \le 4 \cdot \max\{1, \log^2(s_i)\} =: C(s_i).$$

Let A denote the set of indices i that we bound by $\chi(x_i\|s_i)$ and $B = [d]\backslash A$ those that we bound by $C(s_i) \cdot x_i$. We now turn to upper bound the two groups of terms corresponding to these two sets. In particular, let

$$T_1 := \sum_{i\in A} |\chi(p_i\|s_i) - \chi(q_i\|s_i)|,$$

$$T_2 := \sum_{i\in B} |\chi(p_i\|s_i) - \chi(q_i\|s_i)|,$$

respectively, and denote $\Delta_1 := \sum_{i\in A} x_i, \Delta_2 := \sum_{i\in B} x_i$. We can straightforwardly bound T_2 as

$$T_2 \le \sum_{i\in B} C(s_i)x_i \le C(s_{\min})\Delta_2 \le 2C(s_{\min})D,$$

where we recall that $D \equiv D(\rho, \rho')$. Bounding T_1 is more involved. By applying the previously derived upper bound, we have

$$T_1 \le \sum_{i\in A} \chi(x_i\|s_i).$$

We first note

$$\chi(x_i\|s_i) = x_i \log^2(x_i) + 2\eta(x_i)\log(s_i) + x_i \log^2(s_i)$$

$$\le x_i \log^2(x_i) + x_i \log^2(s_i)$$

$$\le x_i \log^2(x_i) + x_i \log^2(x_i e^2)$$

$$= 2x_i \log^2(x_i) - 4\eta(x_i) + 4x_i$$

$$\le 2x_i \log^2(x_i) + 4x_i, \tag{C2}$$

where in the second step we use that $\eta(x_i)$ is positive and $s_i \le 1$, in the second $x_i \le s_i/e^2$, which holds for all the terms in T_1 by the previous arguments, and in the last step again positivity of $\eta(x_i)$. Next, we make use of the identity

$$\log^2(x_i) = \log^2(x_i/\Delta_1) + 2\log(x_i)\log(\Delta_1) - \log^2(\Delta_1).$$

Plugging this into the rhs of Eq. (C2) yields

$$\sum_{i\in A} \chi(x_i\|s_i) \le 2\Delta_1 F_1 - 2F_2 + 4\Delta_1 - \Delta_1 \log^2(\Delta_1), \tag{C3}$$

where

$$F_1 := \sum_{i\in A} \frac{x_i}{\Delta_1} \log^2\left(\frac{x_i}{\Delta_1}\right) = \sum_{i\in A} \chi\left(x_i/\Delta_{\le}\|1\right),$$

$$F_2 := \log(\Delta_1) \sum_{i\in A} \eta(x_i).$$

To bound F_1, we note that $\{x_i/\Delta_{\le}\}_i$ form a $|A|$-dimensional probability vector, corresponding to some density matrix ϱ. Hence,

$$F_1 = L(\varrho\|\mathbb{I}) = V(\varrho\|\mathbb{I}) + S(\varrho\|\mathbb{I})^2$$

$$\le 4\log^2(|A|) \le 4\log^2(d),$$

where we use the fact that, by Property 6 of the variance, as presented in the main text,

$$V(\rho'\|\mathbb{I}) = V(\rho') \le \frac{1}{4}\log^2(d-1) + 1/\ln^2(2) \le 3\log^2(d)$$

and that $S(\varrho\|\mathbb{I})^2 = S(\varrho)^2 \le \log^2(d)$ for a d-dimensional density matrix for the case $|A| \ge 2$. Clearly, this upper bound is also valid for $|A| \in \{0, 1\}$. Next, to lower bound the term F_2, note that

$$\eta(x_i) = \Delta_{\le}\,\eta(x_i/\Delta_{\le}) - x_i\log(\Delta_{\le}),$$

which yields

$$F_2 = \sum_{i\in A} \log(\Delta_1)\left[\Delta_1\,\eta(x_i/\Delta_1) - x_i\log\Delta_1\right]$$

$$\ge \min\{0, \Delta_1\log(\Delta_1)\log(d)\} - \Delta_1\log^2(\Delta_1).$$

The minimization arises because we distinguish two cases: if $\Delta_1 \ge 1$, then the terms $\log(\Delta_1)\Delta_1\,\eta(x_i/\Delta_1)$ are positive and can be lower bounded by zero, while if $\Delta < 1$, the term is negative and can be lower bounded by the $\Delta_1\log(\Delta_1)\log(d)$, again using the fact that the $\{x_i/\Delta_1\}_{i\in A}$ form a probability distribution. Plugging these bounds back into Eq. (C3) then yields

$$T_1 \le 4\Delta_1\left[2\log^2(d) + 4\right]$$

$$+ \chi(\Delta_1\|1) + \max\{0, 2\eta(\Delta_1)\log(d)\}$$

$$\le 8D\left[2\log^2(d) + 4\right] + \chi(\Delta_1\|1)$$

$$+ \max\{0, 2\eta(\Delta_1)\log(d)\},$$

We are finally in a position to combine the bounds on T_1 and T_2. This gives

$$|L(\rho\|\sigma) - L(\rho'\|\sigma)|$$

$$\le 2[C(s_{\min}) + 8\log^2(d) + 8]D$$

$$+ \chi(\Delta_1\|1) + \max\{0, 2\eta(\Delta_1)\log(d)\}$$

$$\le 8[3 + \frac{1}{4}\log(s_{\min})^2 + 2\log^2(d)]D$$

$$+ \chi(\Delta_1\|1) + \max\{0, 2\eta(\Delta_1)\log(d)\},$$

where we use

$$C(s_{\min}) + 8\log^2(d) + 8 = \max\{4, \log^2(s_{\min})\}$$

$$+ 8\log^2(d) + 8 \le 12 + \log^2(s_{\min}) + 8\log^2(d) =: c_1.$$

Now, if $D \le 1/(2e^2)$, then we can use the monotonicity of $\chi(\cdot\|1)$ and η over the interval $[0, 1/e^2]$ to bound $\chi(\Delta_1\|1) \le \chi(2D\|1)$ and $\eta(\Delta_1) \le \eta(2D)$. This provides the first statement of the lemma.

For the second statement, in which we have no promise on the value of the trace distance between ρ and ρ', it suffices to note that, for $x \in [0, 2]$,

$$\eta(x) \le \sqrt{x}, \quad \chi(x\|1) \le 6\sqrt{x}.$$

Since, $\Delta_1 \in [0, 2]$, this implies

$$\chi(\Delta_1\|1) + \max\{0, 2\eta(\Delta_1)\log(d)\}$$
$$\le [e + 2\log(d)]\sqrt{\Delta_1} \le c_2\sqrt{D},$$

where we define $c_2 := [6 + 2\log(d)]$. ∎

With this technical lemma established, it is then relatively easy to prove uniform continuity of $V(\rho\|\sigma)$, which

is stated as Lemma 10 in the main text (we restate it here for convenience).

Lemma 10. (Uniform continuity of the relative variance). *Let σ be a positive-definite operator with $0 < \sigma \le \mathbb{I}$ and smallest eigenvalue $s_{\min}$ on a d-dimensional Hilbert space with $d \ge 2$, and ρ, ρ' be two states that both commute with σ. Then,*

$$|V(\rho\|\sigma) - V(\rho'\|\sigma)| \le 2K\sqrt{D(\rho, \rho')},$$

where $K = 8\log^2(d) + \log(d) + 2\log^2(s_{\min}) - 4\ln(2)\log(s_{\min}) + 15$.

Proof. We have

$$|V(\rho\|\sigma) - V(\rho'\|\sigma)|$$
$$\le |L(\rho\|\sigma) - L(\rho'\|\sigma)| + |S(\rho'\|\sigma)^2 - S(\rho\|\sigma)^2|$$
$$\le |L(\rho\|\sigma) - L(\rho'\|\sigma)| + [S(\rho'\|\sigma) + S(\rho\|\sigma)]|S(\rho'\|\sigma) - S(\rho\|\sigma)|$$
$$\le |L(\rho\|\sigma) - L(\rho'\|\sigma)| - 2\log(s_{\min})|S(\rho'\|\sigma) - S(\rho\|\sigma)|$$
$$\le c_1\epsilon + c_2\sqrt{\epsilon} - 2\log(s_{\min})|S(\rho'\|\sigma) - S(\rho\|\sigma)|$$
$$\le c_1\epsilon + c_2\sqrt{\epsilon} - 2\log(s_{\min})\{(1 + \epsilon)h_b[\epsilon/(1 + \epsilon)] - \log(s_{\min})\epsilon\}$$
$$\le c_1\epsilon + c_2\sqrt{\epsilon} - 2\log(s_{\min})\left[4\ln(2)\sqrt{\epsilon} - \log(s_{\min})\epsilon\right]$$
$$\le \left[c_1 + c_2 - 8\ln(2)\log(s_{\min}) + 2\log^2(s_{\min})\right]\sqrt{\epsilon} =: K\sqrt{\epsilon},$$

where we use $\max\{S(\rho\|\sigma), S(\rho'\|\sigma)\} \le -\log(s_{\min})$ in the third step, Lemma 27 in the fourth step, Lemma 18 in the fifth step, Lemma 16 in the sixth step, and $\epsilon \le \sqrt{\epsilon}$ in the last step (since $\epsilon \in [0, 1]$). ∎

APPENDIX D: PROOF OF CORRECTION TO SUBADDITIVITY OF RELATIVE VARIANCE (LEMMA 11)

For notational convenience, for the remainder of this Appendix we write $V \equiv V(\rho\|\sigma)$ and $V_1 \equiv V(\rho_1\|\sigma_1)$ and similarly for the other subsystem and other quantities, L and S.

Lemma 11. (Correction to subadditivity of relative variance.) *Let ρ, σ be two commuting quantum states on a d-dimensional, bipartite system, with $d \ge 2$ and σ full rank. If $\sigma = \sigma_1 \otimes \sigma_2$ is a product state with smallest eigenvalue*

$s_{\min}$, *then*

$$V(\rho\|\sigma) \le V(\rho_1\|\sigma_1) + V(\rho_2\|\sigma_2) + K'f(I_{1:2}),$$

where $K' = \sqrt{2\ln(2)}\left[12 + \log(s_{\min})^2 + 8\log^2(d)\right]$, $f(x) = \max\{\sqrt[4]{x}, x\}$ *and $I_{1:2}$ denotes the mutual information between the two partitions of ρ.*

Proof. Let us begin by introducing the following shorthand notation:

(a) I as the mutual information of ρ across the partitions 1 and 2;
(b) $D \equiv D(\rho, \rho_1 \otimes \rho_2)$;
(c) $S_\otimes = S(\rho_1 \otimes \rho_2\|\sigma) = S_1 + S_2$;
(d) $L_\otimes = L(\rho_1 \otimes \rho_2\|\sigma) = L_1 + L_2 + 2S_1S_2$.

Next, note that

$$S = I + S_\otimes.$$

The special case of this equality is well known for the non-relative version of von Neumann entropy, i.e., $I(A : B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB})$, while the identity above for relative entropies is proved in Proposition 2 of Ref. [85]. Therefore, we have

$$
\begin{aligned}
V = L - S^2 &= L - L_1 - L_2 + L_1 + L_2 - (I + S_\otimes)^2 \\
&= L - L_1 - L_2 + L_1 + L_2 - I^2 \\
&\quad - 2IS_\otimes - S_1^2 - 2S_1 S_2 - S_2^2 \\
&= V_1 + V_2 + L - L_\otimes - I^2 - 2IS_\otimes \\
&\leq V_1 + V_2 + \zeta_\rho,
\end{aligned}
$$

where $\zeta_\rho := L - L_\otimes$, since in the second last line, the terms I^2 and $2IS_\otimes$ are positive and can be dropped with the inequality. Our goal is to bound this remaining function in terms of the mutual information I. To do so, we can apply Lemma 27, since $[\rho, \sigma] = 0$ implies that $[\rho_1 \otimes \rho_2, \sigma] = 0$. Applying this lemma yields

$$
\begin{aligned}
\zeta_\rho &\leq |L - L_\otimes| \leq c_1 D + c_2 \sqrt{D} \\
&\leq c_1' \sqrt{I} + c_2' \sqrt[4]{I},
\end{aligned}
$$

with $c_1' = \sqrt{2 \ln 2} \cdot c_1$ and $c_2' = \sqrt[4]{2 \ln(2)} c_2$, where c_1, c_2 are the constants from the statement of Lemma 27, and where in the second step we use Pinsker's inequality (Lemma 19) and the fact that $I = S(\rho \| \rho_1 \otimes \rho_2)$. Finally, by noting that, for $d \geq 2$, $c_1' = \max\{c_1', c_2'\}$, and $f(I) = \max\{\sqrt{I}, \sqrt[4]{I}\}$, we obtain the bound

$$
\zeta_\rho \leq c_1' f(I).
$$

The statement of the lemma then follows by setting $K' = c_1'$.   ∎

APPENDIX E: PROOF OF LEMMA 7

In order to proof the central technical result of Lemma 7, we first make the following simple observation of lower and upper bounds for Lorenz curves, which is spelled out in Lemma 28.

Lemma 28. *Given any states ρ, σ such that $[\rho, \sigma] = 0$ and $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, express them in their common eigenbasis $\rho = \sum_i p_i |i\rangle\langle i|$ and $\sigma = \sum_i s_i |i\rangle\langle i|$, where $\{p_i\}$ and $\{s_i\}$ are the eigenvalues of ρ and σ, respectively. Let $l_c(x) = \min(c \cdot x, 1)$ for $x \in [0, 1]$. Denote*

$$
r_{\min} = \min_{i \in P_+} \frac{p_i}{s_i}, \qquad r_{\max} = \max_{i \in P_+} \frac{p_i}{s_i},
$$

where $P_+ = \{i | p_i > 0\}$ is the set of indices for which p_i is strictly positive. Then the Lorenz curve of ρ with respect to

σ *satisfies on the whole interval $x \in [0, 1]$:*

$$
\ell_{r_{\min}}(x) \leq \mathcal{L}_{\rho|\sigma}(x) \leq \ell_{r_{\max}}(x).
$$

Proof. This is obvious given the concavity of the Lorenz curve itself.   ∎

Remark 29: The functions $l_c(x) = \min(c\,x, 1)$ furthermore satisfy the property that whenever $c \geq d$, we have that for all $x \in [0, 1]$, $l_c(x) \geq l_d(x)$.

A small further technical observation stated in Lemma 30 is required to then prove Lemmas 31 and 32, which jointly give rise to Lemma 7.

Lemma 30. *Let ρ, σ be two commuting d-dimensional states that satisfy $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, and denote the ϵ-steep and -flat approximation of ρ with respect to σ as $\rho_{\text{st}}^\epsilon$ and $\rho_{\text{fl}}^\epsilon$, according to Definitions 6 and 5, respectively. Then, for $\epsilon_2 \geq \epsilon_1 \geq 0$,*

$$
\mathcal{L}_{\rho_{\text{st}}^{\epsilon_2}|\sigma}(x) \geq \mathcal{L}_{\rho_{\text{st}}^{\epsilon_1}|\sigma}(x) \geq \mathcal{L}_{\rho|\sigma}(x) \geq \mathcal{L}_{\rho_{\text{fl}}^{\epsilon_1}|\sigma}(x) \geq \mathcal{L}_{\rho_{\text{fl}}^{\epsilon_2}|\sigma}(x).
$$

Proof. The second and third inequality are special cases of the first and fourth inequality, since $\epsilon_1 \geq 0$. Let us first consider the fourth inequality. It is shown in Ref. [24] that the ϵ-flat construction in Definition 5 is the unique state within an ϵ ball of states around ρ, where for any state $\rho' \in \mathcal{B}^\epsilon(\rho)$,

$$
\mathcal{L}_{\rho'|\sigma}(x) \geq \mathcal{L}_{\rho_{\text{fl}}^\epsilon|\sigma}(x), \qquad \forall x \in [0, 1].
$$

Since $\epsilon_2 \geq \epsilon_1$ implies that $\rho_{\text{fl}}^{\epsilon_1} \in \mathcal{B}^{\epsilon_2}(\rho)$, the fourth inequality holds. Lastly, consider the first inequality. First, note that $\rho_{\text{st}}^{\epsilon_1}$ and $\rho_{\text{st}}^{\epsilon_2}$ always share the same basis, so we may write them as $\rho_{\text{st}}^{\epsilon_1} = \sum_{i=1}^d \hat{p}_i^{(1)} |i\rangle\langle i|$ and $\rho_{\text{st}}^{\epsilon_2} = \sum_{i=1}^d \hat{p}_i^{(2)} |i\rangle\langle i|$, respectively. Furthermore, they have the same relative ordering with respect to σ. In other words, the discrete points that define the respective Lorenz curves are aligned with respect to the x axis, and therefore their condition reduces to a simple comparison between the cummulative sum of the eigenvalues. Concretely, we want that for all $k \in \{1, \ldots, d\}$:

$$
\sum_{i=1}^k \hat{p}_i^{(2)} \geq \sum_{i=1}^k \hat{p}_i^{(1)}.
$$

Denoting R_1 and R_2 to be the respective indices R according to the construction of Definition 6, using ϵ_1 and ϵ_2, respectively, note that $R_2 \leq R_1$. For the various regimes

for the Lorenz curves we therefore have

$$k = 1 \qquad \hat{p}_1^{(2)} = p_1 + \epsilon_2 \geq p_1 + \epsilon_1 = \hat{p}_1^{(1)},$$

$$1 < k \leq R_2 \qquad \sum_{i=1}^{k} \hat{p}_i^{(2)} = \epsilon_2 + \sum_{i=1}^{k} p_i \geq \epsilon_1$$

$$+ \sum_{i=1}^{k} p_i = \sum_{i=1}^{k} \hat{p}_i^{(1)},$$

$$k > R_2 \qquad \sum_{i=1}^{k} \hat{p}_i^{(2)} = 1 \geq \sum_{i=1}^{k} \hat{p}_i^{(1)}.$$

This finishes the proof. ∎

Lemma 31. *Let $\epsilon \in [0, 1]$ and let ρ, σ be two commuting d-dimensional states that satisfy $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$, and denote the ϵ-steep approximation of ρ with respect to σ as $\rho_{\mathrm{st}}^{\epsilon}$ according to Definition 6. Then,*

$$\mathcal{L}_{\rho_{\mathrm{st}}^{\epsilon} | \sigma}(x) \geq \ell_{r_{\mathrm{st}}}(x), \quad r_{\mathrm{st}} = 2^{S(\rho \| \sigma) - f_\sigma(\rho, \epsilon)},$$

where $f_\sigma(\rho, \epsilon) := \sqrt{V(\rho \| \sigma) (\epsilon^{-1} - 1)}$ *and* $l_c(x) = \min(cx, 1)$.

Proof. Using the fact that $[\rho, \sigma] = 0$, we can decompose the states into their simultaneous eigenbasis as

$$\rho = \sum_i p_i |i\rangle\langle i|, \quad \sigma = \sum_i s_i |i\rangle\langle i|,$$

and take the ordering of the eigenbasis such that $p_i/s_i \geq p_{i+1}/s_{i+1}$ for all $i \in \{1, \ldots, d-1\}$. Next, given ϵ, define

$$\tilde{\imath} = \max_i \left\{ i \,\middle|\, \sum_{j=i}^{d} p_i \geq \epsilon \right\},$$

namely $\tilde{\imath}$ is the largest index such that the tail sum of the ordered distribution on p is larger or equal to ϵ. Also, denote the following tail sums:

$$\epsilon^+ = \sum_{j=\tilde{\imath}}^{d} p_j, \quad \epsilon^- = \sum_{j=\tilde{\imath}+1}^{d} p_j,$$

where we set $\epsilon^- = 0$ if $\tilde{\imath} = d$. By construction, $\epsilon^- \leq \epsilon \leq \epsilon^+$. We are now going to make use of the steep state with the smaller parameter ϵ^-: let $\rho_{\mathrm{st}}^{\epsilon^-} = \sum_i \hat{p}_i |i\rangle\langle i|$ denote the ϵ^--steep approximation of ρ relative to σ, with $\hat{p}_i$

explicitly defined in Definition 6. By construction, we have that

$$A := \min_{i \in P_+} \left(\frac{\hat{p}_i}{s_i} \right) = \frac{p_{\tilde{\imath}}}{s_{\tilde{\imath}}},$$

where $P_+ = \{i | \hat{p}_i > 0\}$. We can now infer that

$$\mathcal{L}_{\rho_{\mathrm{st}}^{\epsilon} | \sigma}(x) \geq \mathcal{L}_{\rho_{\mathrm{st}}^{\epsilon^-} | \sigma}(x) \geq \ell_A(x),$$

where the first inequality follows from Lemma 30 together with the fact that $\epsilon \geq \epsilon^-$, while the second inequality follows from Lemma 28. Hence, it remains to show that

$$A \geq 2^{S(\rho \| \sigma) - f_\sigma(\rho, \epsilon)}, \tag{E1}$$

which implies $S(\rho \| \sigma) - f_\sigma(\rho, \epsilon) \leq \log A$. If $S(\rho \| \sigma) \leq \log A$, then this inequality would hold for any non-negative function $f_\sigma(\rho, \epsilon)$. Otherwise, we can derive the explicit form of f_σ so that Eq. (E1) holds via Cantelli's inequality. More precisely, consider the real-valued random variable X with sample space $\Omega = \{1, 2, \ldots, d\}$ distributed as $\mathrm{Prob}[X = \log(p_i/s_i)] = p_i$. We then have

$$\epsilon \leq \epsilon^+ = \mathrm{Prob}\,(X \leq \log A)$$

$$= \mathrm{Prob}\,[S(\rho \| \sigma) - X \geq S(\rho \| \sigma) - \log A]$$

$$\leq \frac{V(\rho \| \sigma)}{V(\rho \| \sigma) + [S(\rho \| \sigma) - \log A]^2},$$

where the first step follows by definition of ϵ^+. The second step expresses the fact that, by virtue of the fact that we order the state bases by decreasing ratios p_i/s_i, ϵ^+ is the total probability, with respect to X, that a ratio smaller than or equal to $p_{\tilde{\imath}}/s_{\tilde{\imath}}$ is sampled. In the final step we use Cantelli's inequality (Lemma 21) with random variable $-X$ and $\lambda \equiv S(\rho \| \sigma) - \log A$, together with the fact that the mean and variance of $-X$ are given by $-S(\rho \| \sigma)$ and $V(\rho \| \sigma)$, respectively. The claim then follows by a simple rearrangement of the terms above. ∎

Lemma 32. *Let $\epsilon \in [0, 1]$ and let ρ, σ be two commuting d-dimensional states that satisfy $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$, and denote the ϵ-flat approximation of ρ with respect to σ as $\rho_{\mathrm{fl}}^{\epsilon}$ according to Definition 5. Then,*

$$\mathcal{L}_{\rho_{\mathrm{fl}}^{\epsilon} | \sigma}(x) \leq \ell_{r_{\mathrm{fl}}}(x), \quad r_{\mathrm{fl}} = 2^{S(\rho \| \sigma) + f_\sigma(\rho, \epsilon)},$$

where $f_\sigma(\rho, \epsilon) := \sqrt{V(\rho \| \sigma) (\epsilon^{-1} - 1)}$ *and* $l_c(x) = \min(cx, 1)$.

Proof. The proof is similar in structure to Lemma 31, by using the flat approximation instead of the steep one. We begin as well by writing $\rho = \sum_i p_i |i\rangle\langle i|$ and $\sigma = \sum_i s_i |i\rangle\langle i|$

such that $p_i/s_i \geq p_{i+1}/s_{i+1}$ for all $i \in \{1, \ldots, d-1\}$. Given ϵ, define

$$\tilde{\iota} = \min_i \left\{ i \,\middle|\, \sum_{j=1}^i p_j \geq \epsilon \right\},$$

and set $\epsilon^+ = \sum_{j=1}^{\tilde{\iota}} p_j$. Further, set

$$\epsilon^- = \sum_{j=1}^{\tilde{\iota}-1} p_j - \frac{p_{\tilde{\iota}}}{s_{\tilde{\iota}}} \sum_{j=1}^{\tilde{\iota}-1} s_j,$$

if $\tilde{\iota} > 1$ or $\epsilon^- = 0$ if $\tilde{\iota} = 1$. In either case, we by construction have

$$\epsilon^- \leq \sum_{j=1}^{\tilde{\iota}-1} p_j \leq \epsilon \leq \epsilon^+.$$

Now, let $\rho_{\mathrm{fl}}^{\epsilon^-} = \sum_i \bar{p}_i |i\rangle\langle i|$ denote the ϵ^--flat approximation to ρ relative to σ. By definition of the flat approximation, Definition 5, we can see that for our choice of ϵ^-, we have

$$\frac{\bar{p}_i}{s_i} = \frac{p_{\tilde{\iota}}}{s_{\tilde{\iota}}} \geq \frac{\bar{p}_j}{s_j}, \quad i = 1, \ldots, \tilde{\iota}, \quad j = \tilde{\iota}+1, \ldots, d.$$

To see the left equality, note that we define $\tilde{\iota}$ and ϵ^- in such a way that, in terms of the notation of Definition 5, $M = \tilde{\iota} - 1$. Even for the special case of $\tilde{\iota} = 1$, the above ratio is well defined. Moreover, the definition of the values $\bar{p}_i$ ensures equality of the ratios $\bar{p}_i/s_i$ for $i \leq M$. The right inequality, on the other hand, is a property of the flat approximation that is proven in Ref. [24]. Together, they imply that

$$\max_{l \in [d]} \frac{\bar{p}_l}{s_l} = \frac{p_{\tilde{\iota}}}{s_{\tilde{\iota}}} =: B.$$

Therefore, by Lemma 30, Lemma 28, and Remark 29, we know that $\mathcal{L}_{\rho_{\mathrm{fl}}^\epsilon|\sigma}(x) \leq \mathcal{L}_{\rho_{\mathrm{fl}}^{\epsilon^-}|\sigma}(x) \leq \ell_B(x)$. Our goal is then to show that $B \leq 2^{S(\rho\|\sigma)+f_\sigma(\rho,\epsilon)}$. By positivity of $f_\sigma(\rho,\epsilon)$, this is clearly true whenever $S(\rho\|\sigma) \geq \log B$. In case $S(\rho\|\sigma) < \log B$, we again consider a real-valued random variable X with sample space $\Omega = \{1, 2, \ldots, d\}$ distributed as $\mathrm{Prob}[X = \log(p_i/s_i)] = p_i$. We then have

$$\epsilon \leq \epsilon^+ = \mathrm{Pr}(X \geq \log B)$$
$$= \mathrm{Pr}[X - S(\rho\|\sigma) \geq \log B - S(\rho\|\sigma)]$$
$$\leq \frac{V(\rho\|\sigma)}{V(\rho\|\sigma) + [\log B - S(\rho\|\sigma)]^2},$$

where we use Cantelli's inequality with X and $\lambda \equiv \log(B) - S(\rho\|\sigma) > 0$, in the last step. The claim then follows by rearranging the terms in the above inequality. ∎

APPENDIX F: FINITE BUT LARGE IID SEQUENCES

In this section, we apply Theorem 8 to study sufficient conditions for approximate state transitions in the case of n IID systems. We focus on the case of σ majorization, where $\sigma' = \sigma$. We are interested in the regime where n is large but finite and the error ϵ_n is constant or goes to zero with n, but fulfills $\epsilon_n n \to \infty$. In technical terms, $\sqrt{\epsilon_n}$ is a *moderate sequence* [54].

Lemma 33. *Let* ρ, ρ' *and* σ *be density matrices satisfying* $[\rho, \sigma] = [\rho', \sigma] = 0$, *and that* $\mathrm{supp}(\rho), \mathrm{supp}(\rho') \subseteq \mathrm{supp}(\sigma)$. *Denote* $S \equiv S(\rho\|\sigma), S' \equiv S(\rho'\|\sigma)$ *and* $V \equiv V(\rho\|\sigma), V' \equiv V(\rho'\|\sigma)$, *respectively. Let* $\epsilon_n > 0$ *be a sequence of errors such that* $\epsilon_n n \to \infty$. *Then*

$$\rho^{\otimes n} \succ_{\sigma^{\otimes n}, \epsilon_n} \rho'^{\otimes Rn}, \tag{F1}$$

with rate

$$R \geq \frac{S}{S'} - \sqrt{\frac{2-\epsilon_n}{\epsilon_n n}} g(S, S', V, V', \epsilon_n n) + O\left(\frac{1}{\epsilon_n n}\right),$$

where

$$g(S, S', V, V', \epsilon_n n) := \frac{\sqrt{V}}{S'} - \frac{\sqrt{rV'}}{S'} \sqrt{1 + O(1/\sqrt{\epsilon_n n})}.$$

Proof. According to Theorem 8, the transition in Eq. (F1) is possible as long as

$$S - \sqrt{\frac{V(2-\epsilon_n)}{n\epsilon_n}} > RS' + \sqrt{\frac{RV'(2-\epsilon_n)}{\epsilon_n n}}.$$

Rewriting the above equation in terms of $M = \sqrt{R}$, and dividing throughout by nS', while grouping the terms

$$k = \frac{2-\epsilon_n}{\epsilon_n n} \frac{V}{S'^2}, \qquad k' = \frac{2-\epsilon_n}{\epsilon_n n} \frac{V'}{S'^2}, \tag{F2}$$

the above condition simplifies to

$$M^2 + \sqrt{k'}M - \left[r - \sqrt{k}\right] < 0,$$

where we denote $r = S/S'$ to be the ratio of entropies for the initial and final states. The two roots of this equation give a region $M \in [M^-, M^+]$ for which state transitions may occur. Since we are interested in a sufficient criteria, the lower bound given by M^- is mainly of interest. Solving

the quadratic equation, we then have

$$M > M^- = \frac{1}{2}\left[-\sqrt{k'} - \sqrt{k' + 4(r - \sqrt{k})}\right].$$

Switching back to $R = M^2$, we obtain

$$R > \frac{1}{4}\left[2k' + 4(r - \sqrt{k}) + 2\sqrt{k'}\sqrt{k' + 4(r - \sqrt{k})}\right]$$

$$> r - \sqrt{k} + \sqrt{rk'}\sqrt{1 + O(1/\sqrt{\epsilon_n n})} + O(1/\epsilon_n n),$$

where in the third term, a factor of $O(1/\epsilon_n n)$ has been absorbed into $O(1/\sqrt{\epsilon_n n})$. Recalling the definitions of k, k' concludes the proof. ∎

A result related to Lemma 33 has been derived in Refs. [55,56] for the case of σ being thermal states. In the large n limit, the second-order correction term that has a $1/\sqrt{\epsilon_n n}$ dependence vanishes whenever $k = rk'$ for k, k' defined in Eq. (F2). This indicates that a "resonance" happens whenever

$$\frac{V/S}{V'/S'} = 1,$$

which was observed before in Refs. [55,56]. We should stress, however, that Lemma 33 is far from providing optimal moderate-deviation bounds for the general state-interconversion problem [55,56]. This is to be expected, since we use a single-shot result that is not tailored to the particular structures appearing in the IID limit for large n. However, the above analysis shows that qualitative features may be recovered in a very simple manner by making use of Theorem 8.

APPENDIX G: PROOF OF THEOREM 12

Before moving to the proof, let us first prove the following auxiliary lemma.

Lemma 34. *Let $x, s \in \mathbb{R}^d$ be d-dimensional row vectors with $s > 0$, and $y, s' \in \mathbb{R}^{d'}$ with $s' > 0$, respectively. Furthermore, let A be a $d \times d'$ right stochastic matrix, that is, all entries of A are non-negative and every row sums to 1. If $y = xA$ and $s' = sA$, then*

$$F(y\|s') \geq F(x\|s)$$

for any function of the form $F(x\|s) = \sum_i s_i g\,(x_i/s_i)$ where g is a function that is concave over the interval $[\min_i x_i/s_i, \max_i x_i/s_i]$.

Proof. This proposition is almost the same as stated in Theorem 24, except that in Theorem 24 it holds when the function g is concave over the entire domain. Nevertheless, one should note that it is sufficient for g to be concave over

the entire interval where it is evaluated. This is because given such a g, one can always construct a continuously differentiable function g' such that $g' = g$ within the said interval, and linear outside the interval. Then, g' is concave. Furthermore, since y is related to x by $y = xA$, the maximal and minimal (strictly positive) ratios of the two states are contractive under classical channels. In other words, the smallest and largest relative eigenvalues satisfy $\min_i y_i/s'_i \geq \min_i x_i/s_i$, and $\max_i y_i/s'_i \leq \max_i x_i/s_i$. This follows from Theorem 24 and can be seen, for example, by noting that the maximal and minimal ratios correspond to the initial and final slopes of the Lorenz curves, respectively, so that the above inequalities are implied between Lorenz curves, as stated in that theorem. Therefore, it suffices to have the requirement that g is concave over $[\min_i x_i/s_i, \max_i x_i/s_i]$. ∎

We are now in a position to prove Theorem 12.

Theorem 12. *Let (ρ, σ) and (ρ', σ') be two pairs of commuting quantum states, with σ, σ' both full rank. If $(\rho, \sigma) \succ (\rho', \sigma')$, then it holds that $M_{s_{\min}}(\rho'\|\sigma') \geq M_{s_{\min}}(\rho\|\sigma)$, where*

$$M_x(\rho\|\sigma) := V(\rho\|\sigma) + \left(\frac{1}{\ln(2)} - \log(x) - S(\rho\|\sigma)\right)^2,$$

and $s_{\min}$ denotes the smallest eigenvalue of σ.

Proof. Since $[\rho, \sigma] = 0$, we can express the value $M_{s_{\min}}(\rho\|\sigma)$ in terms of the eigenvalues of $\rho = \sum_i p_i |i\rangle\langle i|$ and $\sigma = \sum_i s_i |i\rangle\langle i|$. Writing $a = 1/\ln(2) - \ln(s_{\min})$, this yields

$$M_{s_{\min}}(\rho\|\sigma) = V(\rho\|\sigma) - 2aS(\rho\|\sigma) + S(\rho\|\sigma)^2 + a^2$$

$$= \operatorname{tr}\left[\rho \log^2\left(\frac{\rho}{\sigma}\right)\right] - 2a\operatorname{tr}\left[\rho \log\left(\frac{\rho}{\sigma}\right)\right] + a^2$$

$$= \sum_i p_i\left[\log^2\left(\frac{p_i}{s_i}\right) - 2a\log\frac{p_i}{s_i}\right] + a^2$$

$$= \sum_i s_i f_a\left(\frac{p_i}{s_i}\right) + a^2 =: F(p\|s) + a^2,$$

where $f_a(x) := x \cdot \left[\log^2(x) - 2a\log(x)\right]$ and $p = (p_1, \ldots, p_d)$ and $s = (s_1, \ldots, s_d)$. By the assumption that $(\rho, \sigma) \succ (\rho', \sigma')$, there exists a quantum channel that maps ρ to ρ' and σ to σ'. Due to the commutation structure, by Lemma 20, this implies that there exists a right stochastic matrix E such that $pE = p'$, with $(p')^T = (p'_1 \ldots, p'_d)$, and that $sE = s'$. We can therefore apply Lemma 34 to find that $F(p\|s) \leq F(p'\|s')$, whenever f_a is concave over the interval $[\min_i p_i/s_i, \max_i p_i/s_i]$. Here, $F(p'\|s')$ is related to $M_{s_{\min}}(\rho'\|\sigma')$ in just the same way as presented for the initial states above, but now using a corresponding common

eigenbasis for ρ' and σ'. Now, it is straightforward to check that for our choice of $a = 1/\ln(2) - \log(s_{\min})$, f_a is concave over the interval $[0, 1/s_{\min}] \supseteq [\min_i p_i/s_i, \max_i p_i/s_i]$, since the second derivative $f_a'' \leq 0$. This establishes that

$$M_{s_{\min}}(\rho\|\sigma) \leq M_{s_{\min}}(\rho'\|\sigma').$$

■

Remark: Theorem 12 is particularly interesting because it hints at the possibility of constructing a new classical entropic measure (as a function of relative entropy and variance) that satisfies the data-processing inequality. Note that this is not yet completely true in Theorem 12 because the monotone $M_{s_{\min}}$ is dependent on the initial reference state.

APPENDIX H: PROOF OF RESULT 5 AND THEOREM 15

In this Appendix, we prove that local monotonicity singles out von Neumann entropy and the relative entropy among continuous functions on (pairs of) quantum states. We first develop properties of functions that apply in both settings and then prove the respective results. In both cases, we can view the function f in question as acting on pairs of quantum states over certain subsets of density matrices, since in the case of local monotonicity with respect to maximally mixed states we can simply view $f(\rho)$ as $f(\rho,\sigma)$ with σ being the maximally mixed state of the same dimension as ρ. More generally, let us consider a set of density matrices $\mathcal{S}$ that is

1. Closed under tensor products: $\sigma, \sigma' \in \mathcal{S} \Rightarrow \sigma \otimes \sigma' \in \mathcal{S}$.
2. Closed under permutations: $\sigma \otimes \sigma' \in \mathcal{S} \Rightarrow \sigma' \otimes \sigma \in \mathcal{S}$.

Furthermore, for any $\sigma \in \mathcal{S}$, denote by $\mathcal{C}_\sigma$ the set of channels that leave the state invariant:

$$C \in \mathcal{C}_\sigma \Rightarrow C[\sigma] = \sigma,$$

and denote by $\mathcal{H}_\sigma$ the Hilbert space on which $\sigma \in \mathcal{S}$ is defined. Suppose $\sigma_1, \sigma_2 \in \mathcal{S}$, and $C \in \mathcal{C}_{\sigma_1 \otimes \sigma_2}$ is given and consider two states ρ_i on $\mathcal{H}_{\sigma_i}$ for $i = 1, 2$. Then we write

$$\rho_1' := \mathrm{tr}_2[C(\rho_1 \otimes \rho_2)], \quad \rho_2' := \mathrm{tr}_1[C(\rho_1 \otimes \rho_2)].$$

A function f on pairs of quantum states (ρ, σ) with $\sigma \in \mathcal{S}$ and $\rho \in \mathcal{D}(\mathcal{H}_\sigma)$ is called *locally monotonic with respect to $\mathcal{S}$* if

$$f(\rho_1, \sigma_1) + f(\rho_2, \sigma_2) \geq f(\rho_1', \sigma_1) + f(\rho_2', \sigma_2) \quad \text{(H1)}$$

for any such pairs (ρ_i, σ_i) and channels $C \in \mathcal{C}_{\sigma_1 \otimes \sigma_2}$. We here use the $\geq$ sign in the definition, since we are interested in results relative to the states in $\mathcal{S}$.

We can further generalize the definition of local monotonicity. Let $\mathcal{C}_\mathcal{S}$ be the set of channels that map states from $\mathcal{S}$ to $\mathcal{S}$ (provided they are in the domain of the corresponding channel). Then we say that f is *locally monotonic with respect to $\mathcal{C}_\mathcal{S}$* if for all channels $C \in \mathcal{C}_\mathcal{S}$ and $\sigma_1, \sigma_2 \in \mathcal{S}$ such that $C(\sigma_1 \otimes \sigma_2) = \sigma_1' \otimes \sigma_2' \in \mathcal{S}$ we have

$$f(\rho_1, \sigma_1) + f(\rho_2, \sigma_2) \geq f(\rho_1', \sigma_1') + f(\rho_2', \sigma_2')$$

if ρ_1 and ρ_2 are density matrices on the respective Hilbert space associated to σ_1 and σ_2. A function that is locally monotonic with respect to $\mathcal{C}_\mathcal{S}$ is always locally monotonic with respect to $\mathcal{S}$. In the following, we therefore first prove general properties of functions that are locally monotonic with respect to $\mathcal{S}$.

Given a function f that is locally monotonic with respect to $\mathcal{S}$, we define a function f' as

$$f'(\rho, \sigma) := f(\rho, \sigma) - f(\sigma, \sigma).$$

Then f' is still locally monotonic with respect to $\mathcal{S}$, since the terms of the form $f(\sigma, \sigma)$ cancel in the corresponding Eq. (H1).

Lemma 35. *If f is locally monotonic with respect to $\mathcal{S}$, then $f(\rho, \sigma) \geq f(C[\rho], \sigma)$ for any $C \in \mathcal{C}_\sigma$ and the same is true for f'.*

Proof. If $C \in \mathcal{C}_\sigma$, then $C \otimes \mathbf{1} \in \mathcal{C}_{\sigma \otimes \sigma}$. But then, since $C[\sigma] = \sigma$, we have

$$f(\rho, \sigma) + f(\sigma, \sigma) \geq f[C(\sigma), \sigma] + f(\sigma, \sigma).$$

Since $f'(\rho, \sigma) - f'(C[\rho], \sigma) = f(\rho, \sigma) - f(C[\rho], \sigma)$, the same is true for f'. ■

Lemma 36. *If f is locally monotonic with respect to $\mathcal{S}$, then f' is additive under tensor products:*

$$f'(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) = f'(\rho_1, \sigma_1) + f'(\rho_2, \sigma_2).$$

Proof. Consider the pairs $(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2)$, (σ_1, σ_1) and the channel $C \in \mathcal{C}_{(\sigma_1 \otimes \sigma_2) \otimes \sigma_1}$ that permutes the first subsystem of the first pair with the second system. We then find

$$f'(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) + f'(\sigma_1, \sigma_1) \geq f'(\sigma_1 \otimes \sigma_2, \sigma_1 \otimes \sigma_2) + f'(\rho_1, \sigma_1).$$

But since $f'(\sigma_1, \sigma_1) = 0$, we find $f'(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) \geq f'(\rho_1, \sigma_1)$. The permutation channel is reversible and by

considering the reverse transition, we find the converse relation. Hence

$$f'(\rho_1 \otimes \sigma_2, \sigma_1 \otimes \sigma_2) = f'(\rho_1, \sigma_1), \quad \forall \sigma_2 \in \mathcal{S}.$$

Similarly, we get

$$f'(\sigma_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) = f'(\rho_2, \sigma_2), \quad \forall \sigma_1 \in \mathcal{S}.$$

Considering now the pairs $(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2)$ and (σ_1, σ_1), we similarly find

$$f'(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) + \underbrace{f'(\sigma_1, \sigma_1)}_{=0}$$
$$= f'(\sigma_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) + f'(\rho_1, \sigma_1)$$
$$= f'(\rho_1, \sigma_1) + f'(\rho_2 \otimes \sigma_2).$$

We thus find that f' is additive under tensor products. ∎

Lemma 37. *If f is locally monotonic with respect to $\mathcal{S}$, then f' is superadditive:*

$$f'(\rho_{12}, \sigma_1 \otimes \sigma_2) \geq f'(\rho_1, \sigma_1) + f'(\rho_2, \sigma_2).$$

Proof. Consider the pairs $(\rho_{12}, \sigma_1 \otimes \sigma_2)$, (ρ_1, σ_1) and again the channel that swaps the first subsystem of the first pair with the second system as in the proof of the previous lemma. Then

$$f'(\rho_{12}, \sigma_1 \otimes \sigma_2) + f'(\rho_1, \sigma_1)$$
$$\geq f'(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) + f'(\rho_1, \sigma_1)$$
$$= f'(\rho_1, \sigma_1) + f'(\rho_2, \sigma_2) + f'(\rho_1, \sigma_1),$$

where we use additivity of f' in the last line. ∎

To summarize, we have found that if f is locally monotone with respect to $\mathcal{S}$, then

$$f(\rho, \sigma) = f'(\rho, \sigma) + f(\sigma, \sigma),$$

with f' being additive and superadditive over tensor products and monotonic under the channels $\mathcal{C}_\sigma$: $f'(\rho, \sigma) \geq f'(C[\rho], \sigma)$. Result 5 now follows as the following corollary by considering as $\mathcal{S}$ the set of maximally mixed states.

Corollary 38. *Let $\mathcal{S}$ consist of all maximally mixed states and let f be locally monotonic with respect to $\mathcal{S}$ and continuous (for fixed σ). Then*

$$f(\rho) := \log(d) - f(\rho, \mathbb{I}/d) = aS(\rho) + b_d,$$

where a is a constant and b_d a constant that only depends on the Hilbert-space dimension d of ρ.

Proof. In this case, all unitary channels are included in the set of channels. Note that this in particular implies that $f(U\rho U^\dagger, \mathbb{I}/d) = f(\rho, \mathbb{I}/d)$ since unitary channels are reversible. Lemmas 35–37 now show that $g'(\rho) := -f'(\rho, \mathbb{I}/d)$ fulfills the conditions of Lemma 9 in Ref. [78], which shows that

$$-f'(\rho, \mathbb{I}/d) = g'(\rho) = aS(\rho) + b'_d.$$

Since $f'(\mathbb{I}/d, \mathbb{I}/d) = 0$, we have $b'_d = -a \log(d)$. We thus get

$$f(\rho) = \log(d) - [f(\rho, \mathbb{I}/d)' + f(\mathbb{I}/d, \mathbb{I}/d)] = \log(d)$$
$$+ aS(\rho) - a \log(d) - f(\mathbb{I}/d, \mathbb{I}/d) = aS(\rho) + b_d.$$

∎

Let us now consider the setting where $\mathcal{S} = \mathcal{F}$ is the set of finite-dimensional density matrices of full rank. We make use of the following adaption of Lemma 35.

Lemma 39. *If f is locally monotonic with respect to $\mathcal{C}_\mathcal{F}$, then $f(\rho, \sigma) \geq f(C[\rho], C[\sigma])$ for any $C \in \mathcal{C}_\mathcal{F}$ and the same is true for f'. Moreover, $f(\sigma, \sigma)$ is independent of $\sigma \in \mathcal{F}$.*

Proof. If $C_1 \in \mathcal{C}_\mathcal{F}$ and $C_2 \in \mathcal{C}_\mathcal{F}$, then also $C_1 \otimes C_2 \in \mathcal{C}_\mathcal{F}$. In particular, we may choose $C_1 = C$ and $C_2 = \mathbf{1}$ as the identity quantum channel acting on $D(\mathcal{H}_\sigma)$. By local monotonicity for the initial state $(\rho \otimes \sigma, \sigma \otimes \sigma)$ we then have

$$f(\rho, \sigma) + f(\sigma, \sigma) \geq f(C[\rho], C[\sigma]) + f(\sigma, \sigma).$$

Hence $f(\rho, \sigma) \geq f(C[\rho], C[\sigma])$ since by continuity of f, $f(\sigma, \sigma)$ must be finite. But for every two states $\sigma_1, \sigma_2 \in \mathcal{F}$ there are quantum channels in $\mathcal{C}_\mathcal{F}$ mapping one to the other (namely the constant channels). Therefore,

$$f(\sigma_1, \sigma_1) \geq f(\sigma_2, \sigma_2) \geq f(\sigma_1, \sigma_1).$$

Hence $f(\sigma, \sigma)$ is independent of $\sigma \in \mathcal{F}$. This further implies that $f'(\rho, \sigma) - f'(C[\rho], C[\sigma]) = f(\rho, \sigma) - f(C[\rho], C[\sigma])$. ∎

Since the quantum channels that swap Hilbert-space tensor factors do not reduce the rank of any state, they are contained in $\mathcal{C}_F$. Therefore, the proofs of Lemmas 36 and 37 transfer verbatim to the present case. We can now prove Theorem 15, which we restate here for completeness.

Theorem 15. *Let f be a function that is locally monotonic with respect to $\mathcal{C}_\mathcal{F}$ and assume that $\rho \mapsto f(\rho, \sigma)$ is*

continuous for fixed $\sigma \in \mathcal{F}$. Then

$$f(\rho, \sigma) = aS(\rho\|\sigma) + b,$$

where a and b are constants.

Proof of Theorem 15. As before, we define $f'(\rho, \sigma) = f(\rho, \sigma) - f(\sigma, \sigma)$. By Lemma 39, we know that $f(\sigma, \sigma) = b$ is a constant and f' is monotonic under arbitrary quantum channels that map full-rank states to full-rank states (on possibly different Hilbert spaces):

$$f'(\rho, \sigma) \geq f'[C(\rho), C(\sigma)].$$

Furthermore f' is continuous (by assumption), additive, and superadditive (by the same arguments as in Lemmas 36 and 37 as mentioned above). Thus f' is continuous, additive, superadditive, and monotonic under quantum channels mapping full-rank states to full-rank states and hence fulfills the conditions of the main result of Ref. [86]. This implies

$$f'(\rho, \sigma) = aS(\rho\|\sigma).$$

Hence

$$f(\rho, \sigma) = f'(\rho, \sigma) + b = aS(\rho\|\sigma) + b.$$

∎

APPENDIX I: RELATION BETWEEN RÉNYI ENTROPIES AND CUMULANTS OF SURPRISAL

Here, we show the relation between the cumulants of surprisal and the Rényi entropies claimed in the concluding section. For simplicity, we restrict to the unital case, where we have to only deal with entropies instead of divergences. Similar reasoning applies in the case of divergences, though. We first define the *cumulant-generating function $K_X : \mathbb{R} \to \mathbb{R}$* of a real-valued random variable X as

$$K_X(t) := \log_2[\mathbb{E}(2^{tX})],$$

whenever X is such that this function exists. The nth cumulant for a given K_X is defined as $\kappa^{(n)} := K_X^{(n)}(t)|_{t=0}$, that is, as the nth derivative of $K(t)$ evaluated at $t = 0$. If K_X exists, then it is always infinitely differentiable and so all the cumulants are well defined.

In our case, we consider a finite-dimensional, positive density matrix $\rho = \sum_i p_i |i\rangle\langle i| > 0$ and the real random variable $X := -\log(\rho)$ distributed as $\text{Prob}[-\log(\rho) = -\log(p_i)] = p_i$ (by slight abuse of notation). Then the cumulant generating function is given in terms of the Rényi

entropies as

$$K_{-\log(\rho)}(t) = \log_2[\mathbb{E}(2^{-t\log(\rho)})]$$
$$= \log_2(\text{tr}[\rho^{1-t}]) = tS_{1-t}(\rho).$$

In particular, it exists for $t \in (-\infty, 1)$ and the cumulants are well defined.

We then have the following.

Lemma 40. *Let $\rho = \sum_i p_i |i\rangle\langle i| > 0$ be a positive-definite density operator and let $-\log(\rho)$ denote the surprisal with respect to ρ, i.e., the random variable distributed as $\text{Prob}[-\log(\rho) = -\log(p_i)] = p_i$. Then for any $\alpha \in (0, \infty)$,*

$$S_\alpha(\rho) = \sum_{n=1}^{\infty} \frac{\kappa^{(n)}}{n!} (1 - \alpha)^{n-1},$$

where the cumulants are defined with respect to $X = -\log(\rho)$.

Proof. We first note that the Rényi entropies are related to the cumulant-generating function with respect to the surprisal as

$$\kappa := K_{-\log(\rho)}(t) = \log_2[\mathbb{E}(2^{-t\log(\rho)})]$$
$$= \log_2(\text{tr}[\rho^{1-t}]) = tS_{1-t}(\rho),$$

for $t \in [-\infty, 1]$ [for $t \in \{1, 0, -\infty\}$ this follows by continuity of the curve $\alpha \mapsto S_\alpha(\rho)$ [87]]. Using this relation and repeatedly applying the product rule we find

$$\kappa^{(1)} = S_{1-t}(\rho) - tS_{1-t}^{(1)}(\rho),$$
$$\kappa^{(2)} = -2S_{1-t}^{(1)}(\rho) + tS_{1-t}^{(2)}(\rho),$$
$$\dots$$
$$\kappa^{(n)} = (-1)^{n-1}nS_{1-t}^{(n-1)}(\rho) + (-1)^n tS_{1-t}^{(n)}(\rho),$$

where $S_a^{(n)}(\rho)$ denotes the nth derivative of the curve $\alpha \mapsto S_\alpha(\rho)$ evaluated at $\alpha = a$. The derivatives $S_{1-t}^{(n)}$ clearly exist for $t \in (-\infty, 0) \cup (0, 1)$. Lemma 41 shows $S_\alpha(\rho)$ is also infinitely differentiable at $a = 1$ (i.e., $t = 0$), so that the derivatives $S_1^{(n)}$ are well defined and bounded.

At the point $\alpha = 1$ (i.e., $t = 0$), some terms from the above set of equations vanish. Writing out the expression for $\kappa^{(n+1)}$ and taking $\alpha = 1$ (i.e., $t = 0$) yields

$$S_1^{(n)}(\rho) = (-1)^n \frac{\kappa^{(n+1)}}{n+1}.$$

Therefore, the Taylor expansion of $S_\alpha(\rho)$ around $\alpha = 1$ can be rewritten in terms of the cumulants of

surprical as

$$S_\alpha(\rho) = \sum_{n=0}^{\infty} \frac{S_1^{(n)}(\rho)}{n!}(\alpha-1)^n = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}\kappa^{(n)}}{n!}$$

$$\times (\alpha-1)^{n-1} = \sum_{n=1}^{\infty} \frac{\kappa^{(n)}}{n!}(1-\alpha)^{n-1},$$

where as usual we define $0! = 1$ and $0^0 = 1$. $\blacksquare$

What is left, is to show that the function $\alpha \mapsto S_\alpha(\rho)$ is infinitely differentiable at $\alpha = 1$ or, more precisely, can be extended over $\alpha = 1$ analytically. We use complex analysis and denote by $\dot{B}_\epsilon$ an open ball of radius ϵ around $z = 1$ in the complex plane.

Lemma 41. *For any finite-dimensional quantum state ρ, there exists an $\epsilon > 0$ such that the function $z \mapsto S_z(\rho)$ is holomorphically extendable over $z = 1$ on $\dot{B}_\epsilon \subset \mathbb{C}$, i.e., there exists a holomorphic function $g(z)$ on $\dot{B}_\epsilon$ that coincides with $z \mapsto S_z(\rho)$ on $\dot{B}_\epsilon \setminus \{1\}$.*

Proof. We write

$$S_z(\rho) = \frac{1}{1-z}\log Q_z, \quad Q_z = \text{tr}[\rho^z] = \sum_i p_i^z,$$

where the p_i denote the eigenvalues of ρ. The function Q_z is clearly holomorphic. Moreover, for sufficiently small imaginary values of z, it is nonzero. Thus there exists an $\epsilon > 0$ such that Q_z is both holomorphic and nonzero on $\dot{B}_\epsilon$. Therefore, there exists a holomorphic branch of the logarithm on $\dot{B}_\epsilon$, so that $g(z) = \log Q_z$ is holomorphic on $\dot{B}_\epsilon$. Clearly then $S_z(\rho)$ is holomorphic on $\dot{B}_\epsilon \setminus \{1\}$. We also have that

$$\lim_{z\to 1}(z-1)S_z(\rho) = \lim_{z\to 1} -\log Q_z = 0.$$

Therefore, by Riemann's theorem on removable singularities, $S_z(\rho)$ can be holomorphically extended over the point $z = 1$. $\blacksquare$

Note that $g(z)$ being holomorphic, it is in particular continuous, so that $g(1) = \lim_{\alpha\to 1} S_\alpha(\rho) = S(\rho)$. In other words $g(\alpha) = S_\alpha(\rho)$ for all $\alpha \in (1-\epsilon, 1+\epsilon)$.

APPENDIX J: $S_k(\rho)$ FOR $k = 2,\ldots,d$ ENCODE THE SPECTRUM OF ρ

Theorem 42. *The eigenvalues of ρ can be uniquely determined (including multiplicities) from the values $S_k(\rho)$ for $k = 2,\ldots,d$.*

Proof. To our knowledge the proof sketch for this theorem first appeared as a comment by Steve Flammia on the

website mathoverflow [88], which we expand here for the reader's convenience. Let the eigenvalues of ρ be given by p_j with $j = 1,\ldots,d$. Then for $k \geq 2$ we can express the kth *power sum* of the p_j as

$$\sum_j p_j^k = \exp[(1-k)S_k(\rho)].$$

By normalization, we always have $\sum_j p_j = 1$. So only the power sums for $k \geq 2$ provide new information. The power sums can be used to recursively compute the *elementary symmetric polynomials* $e_j(p_1,\ldots,p_d)$ for $j = 0,\ldots,d$ using the Girard-Newton identities [89] as [with $e_0(p_1,\ldots,p_d) = 1$]

$$ke_k(p_1,\ldots,p_d) = \sum_{i=1}^{k}(-1)^{i-1}e_{k-i}(p_1,\ldots,p_d)\sum_j p_j^i$$

$$= \sum_{i=1}^{k}(-1)^{i-1}e_{k-i}(p_1,\ldots,p_d)$$

$$\times \exp[(1-i)S_i(\rho)].$$

Note that only the knowledge of the power sums, or equivalently Rényi entropies, for $k = 2,\ldots,d$ are required to compute the elementary symmetric polynomials. Finally, we can express the characteristic polynomial $c_\rho(\lambda)$ of ρ as a sum over the d elementary symmetric polynomials [89]:

$$c_\rho(\lambda) = \sum_{k=0}^{d}(-1)^k e_k(p_1,\ldots,p_d)\lambda^{d-k}.$$

Solving for the roots of the characteristic polynomial then gives us a unique solution, which is the set of eigenvalues p_j of ρ (including multiplicities). To summarize, we have expressed the characteristic polynomial $C_\rho(\lambda)$ of ρ in terms of the Rényi entropies $S_k(\rho)$ for $k = 2,\ldots,d$ and solving for the roots of the characteristic polynomial allows us to determine the spectrum of ρ. $\blacksquare$

The above result shows that, in principle, the Rényi entropies S_k with $k \geq 2$ uniquely determine the spectrum of a density matrix. In fact the procedure is relatively simple to implement in Wolfram Mathematica. For purely illustrative purposes we therefore include the following code. While it is not particularly numerically stable, it can be seen to work well for small-dimensional matrices by comparing its output with the built-in routine `Eigenvalues[]`.

```
(* Computes k−th power−sums of a square matrix A. For a density matrix A, this is equivalent to defining the following
    function as Exp[(1−k)S_k[A]] for k>= 2 if the k−th Renyi entropy of A is given by S_k[A] and Tr[A] for k=1.*)
PowerSum[k_, A_] := Tr[MatrixPower[A, k]];

(* Computes k−th elementary symmetric polynomial of matrix A from power−sums. *)
e[k_, A_] := If[k == 0, 1,
            1/k Sum[(-1)^(i-1) e[k-i, A] PowerSum[i, A], {i, 1, k}]
        ];

(* Constructs characteristic polynomial from elmentary symmetric polynomials. *)
CharPol[A_] := Module[{d},
                d = Length[A];
                Sum[(-1)^k e[k, A] x^(d-k), {k, 0, d}]
        ];

(* Solve for roots of characteristic polynomial to obtain eigenvalues. Output as list. *)
eigenvalues[A_] := x /. {ToRules[NRoots[CharPol[A] == 0, x]]} // Flatten;
```

[1] B. Schumacher, Quantum coding, Phys. Rev. A **51**, 2738 (1995).

[2] R. Jozsa and B. Schumacher, A new proof of the quantum noiseless coding Theorem, J. Mod. Opt. **41**, 2343 (1994).

[3] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Concentrating partial entanglement by local operations, Phys. Rev. A **53**, 2046 (1996).

[4] B. Schumacher and M. D. Westmoreland, Sending classical information via noisy quantum channels, Phys. Rev. A **56**, 131 (1997).

[5] A. Holevo, The capacity of the quantum channel with general signal states, IEEE Trans. Inf. Theory **44**, 269 (1998).

[6] S. Lloyd, Capacity of the noisy quantum channel, Phys. Rev. A **55**, 1613 (1997).

[7] P. M. Hayden, M. Horodecki, and B. M. Terhal, The asymptotic entanglement cost of preparing a quantum state, J. Phys. A: Math. Gen. **34**, 6891 (2001).

[8] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, Resource Theory of Quantum States Out of Thermal Equilibrium, Phys. Rev. Lett. **111**, 250404 (2013).

[9] M. Horodecki, P. Horodecki, and J. Oppenheim, Reversible transformations from pure to mixed states and the unique measure of information, Phys. Rev. A **67**, 062104 (2003).

[10] W. Hoeffding, The Annals of Mathematical Statistics, 169 (1952).

[11] T. M. Cover, *Elements of Information Theory* (John Wiley & Sons, New York, 1999).

[12] G. H. Hardy, J. E. Littlewood, and G. Pólya, Some simple inequalities satisfied by convex functions. Messenger Math. **58**, 145 (1929).

[13] D. Blackwell, Equivalent comparisons of Experiments, Ann. Math. Statist. **24**, 265 (1953).

[14] E. Ruch, R. Schranner, and T. H. Seligman, The mixing distance, J. Chem. Phys. **69**, 386 (1978).

[15] E. Ruch, R. Schranner, and T. H. Seligman, Generalization of a theorem by Hardy, Littlewood, and Pólya, J. Math. Anal. Appl. **76**, 222 (1980).

[16] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and its Applications* (Springer Science+Business Media, LLC, 2011).

[17] F. Buscemi and G. Gour, Quantum relative Lorenz curves, Phys. Rev. A **95**, 012110 (2017).

[18] X. Wang and M. M. Wilde, Resource theory of asymmetric distinguishability for quantum channels, Phys. Rev. Res. **1**, 033169 (2019).

[19] M. Horodecki and J. Oppenheim, Fundamental limitations for quantum and nanoscale thermodynamics, Nat. Commun. **4**, 1 (2013).

[20] J. M. Renes, Relative submajorization and its use in quantum resource theories, J. Math. Phys. **57**, 122202 (2016).

[21] A. Winter and D. Yang, Operational Resource Theory of Coherence, Phys. Rev. Lett. **116**, 120404 (2016).

[22] G. Gour, Quantum resource theories in the single-shot regime, Phys. Rev. A **95**, 062314 (2017).

[23] E. Chitambar and G. Gour, Quantum resource theories, Rev. Mod. Phys. **91**, 025001 (2019).

[24] R. van der Meer, N. H. Y. Ng, and S. Wehner, Smoothed generalized free energies for thermodynamics, Phys. Rev. A **96**, 062135 (2017).

[25] M. Horodecki, J. Oppenheim, and C. Sparaciari, Extremal distributions under approximate majorization, J. Phys. A: Math. Theor. **51**, 305301 (2018).

[26] R. Renner and S. Wolf, in *International Symposium on Information Theory ISIT* (IEEE, 2004).

[27] R. Renner, *Security of Quantum Key Distribution*, Ph.D. thesis, school ETH Zurich (2005), ArXiv:quant-ph/0512258.

[28] R. Renner and S. Wolf, in *Lecture Notes in Computer Science* (Springer Berlin Heidelberg, 2005), p. 199.

[29] R. Renner, S. Wolf, and J. Wullschleger, in *2006 IEEE International Symposium on Information Theory* (IEEE, 2006).

[30] R. Konig, R. Renner, and C. Schaffner, The operational meaning of min- and max-entropy, IEEE Trans. Inf. Theory **55**, 4337 (2009).

[31] F. Buscemi and N. Datta, The quantum capacity of channels with arbitrarily correlated noise, IEEE Trans. Inf. Theory **56**, 1447 (2010).

[32] M. Tomamichel, R. Colbeck, and R. Renner, Duality between smooth min- and max-entropies, IEEE Trans. Inf. Theory **56**, 4674 (2010).

[33] F. G. S. L. Brandao and N. Datta, One-shot rates for entanglement manipulation under non-entangling maps, IEEE Trans. Inf. Theory **57**, 1754 (2011).

[34] L. Wang and R. Renner, One-Shot Classical-Quantum Capacity and Hypothesis Testing, Phys. Rev. Lett. **108**, 200501 (2013).

[35] N. Datta, J. M. Renes, R. Renner, and M. M. Wilde, One-shot lossy quantum data compression, IEEE Trans. Inf. Theory **59**, 8057 (2013).

[36] N. Datta, M. Mosonyi, M.-H. Hsieh, and F. G. S. L. Brandao, A smooth entropy approach to quantum hypothesis testing and the classical capacity of quantum channels, IEEE Trans. Inf. Theory **59**, 8014 (2013).

[37] F. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, The second laws of quantum thermodynamics, Proc. National Acad. Sci. **112**, 3275 (2015).

[38] M. Tomamichel, *Quantum Information Processing with Finite Resources* (Springer International Publishing, Cham, 2016).

[39] A. Anshu, V. K. Devabathini, and R. Jain, Quantum Communication Using Coherent Rejection Sampling, Phys. Rev. Lett. **119**, 120506 (2017).

[40] V. Strassen, Transactions of the Third Prague Conference on Information Theory etc, 1962. Czechoslovak Academy of Sciences, Prague, 689 (1962).

[41] M. Hayashi, Second-order asymptotics in fixed-length source coding and intrinsic randomness, IEEE Trans. Inf. Theory **54**, 4619 (2008).

[42] M. Hayashi, Information spectrum approach to second-order coding rate in channel coding, IEEE Trans. Inf. Theory **55**, 4947 (2009).

[43] Y. Polyanskiy, H. V. Poor, and S. Verdu, Channel coding rate in the finite blocklength regime, IEEE Trans. Inf. Theory **56**, 2307 (2010).

[44] M. Tomamichel and M. Hayashi, A hierarchy of information quantities for finite block length analysis of quantum tasks, IEEE Trans. Inf. Theory **59**, 7693 (2013).

[45] S. Verdu and I. Kontoyiannis, in *2012 46th Annual Conference on Information Sciences and Systems (CISS)* (IEEE, 2012).

[46] Y. Altug and A. B. Wagner, Moderate deviations in channel coding, IEEE Trans. Inf. Theory **60**, 4417 (2014).

[47] K. Li, Second-order asymptotics for quantum hypothesis testing, Ann. Statistics **42**, 171 (2014).

[48] N. Datta and F. Leditzky, Second-order asymptotics for source coding, dense coding, and pure-state entanglement conversions, IEEE Trans. Inf. Theory **61**, 582 (2014).

[49] V. Y. F. Tan, Asymptotic estimates in information theory with non-vanishing error probabilities, Found. Trends Commun. Inf. Theory **11**, 1, (2014).

[50] M. Tomamichel and V. Y. F. Tan, Second-order asymptotics for the classical capacity of image-additive quantum Channels, Commun. Math. Phys. **338**, 103 (2015).

[51] M. Tomamichel, M. Berta, and J. M. Renes, Quantum coding with finite resources, Nat. Commun. **7**, 11419 (2016).

[52] W. Kumagai and M. Hayashi, Second-order asymptotics of conversions of distributions and entangled states based on rayleigh-normal probability distributions, IEEE Trans. Inf. Theory **63**, 1829 (2016).

[53] C. T. Chubb, M. Tomamichel, and K. Korzekwa, Quantum **2**, 108 (2018).

[54] C. T. Chubb, V. Y. F. Tan, and M. Tomamichel, Moderate deviation analysis for classical communication over quantum Channels, Commun. Math. Phys. **355**, 1283 (2017).

[55] C. T. Chubb, M. Tomamichel, and K. Korzekwa, Moderate deviation analysis of majorization-based resource interconversion, Phys. Rev. A **99**, 032332 (2019).

[56] K. Korzekwa, C. T. Chubb, and M. Tomamichel, Avoiding Irreversibility: Engineering Resonant Conversions of Quantum Resources, Phys. Rev. Lett. **122**, 110403 (2019).

[57] P. Alberti and A. Uhlmann, A problem relating to positive linear maps on matrix algebras, Rep. Math. Phys. **18**, 163 (1980).

[58] F. Buscemi, D. Sutter, and M. Tomamichel, An information-theoretic treatment of quantum dichotomies, Quantum **3**, 209 (2019).

[59] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Yunger Halpern, The resource theory of informational nonequilibrium in thermodynamics, Phys. Rep. **583**, 1 (2015).

[60] In general, unital channels are channels that map the identity of their domain to the identity of their image, and hence are more general than the channels we consider here. We ignore this difference because here we are interested only in strict preservation of the input state.

[61] There exist various equivalent definitions of majorization between quantum states. At the level of d-dimensional probability distributions, we say that $\vec{p} \succ_{\mathbb{I}} \vec{q}$ if and only if for all $k = 1, 2, \ldots, n - 1$ it holds that $\sum_{i=1}^{k} p_i^{\downarrow} \geq \sum_{i=1}^{k} q_i^{\downarrow}$. An alternative definition to the one given above is to say that ρ majorizes ρ' when the vector of eigenvalues of ρ majorizes that of ρ'.

[62] H. Yao and X.-L. Qi, Entanglement Entropy and Entanglement Spectrum of the Kitaev Model, Phys. Rev. Lett. **105**, 080501 (2010).

[63] J. Schliemann, Entanglement spectrum and entanglement thermodynamics of quantum Hall bilayers at v=1, Phys. Rev. B **83**, 115322 (2011).

[64] J. de Boer, J. Järvelä, and E. Keski-Vakkuri, Aspects of capacity of entanglement, Phys. Rev. D **99**, 066012 (2019).

[65] F. Dupuis and O. Fawzi, Entropy accumulation with improved second-order term, IEEE Trans. Inf. Theory **65**, 7596 (2019).

[66] D. Reeb and M. M. Wolf, Tight bound on relative entropy by entropy difference, IEEE Trans. Inf. Theory **61**, 1458 (2015).

[67] P. Faist, T. Sagawa, K. Kato, H. Nagaoka, and F. G. Brandão, Macroscopic Thermodynamic Reversibility in Quantum Many-Body Systems, Phys. Rev. Lett. **123**, 250601 (2019).

[68] T. Sagawa, P. Faist, K. Kato, K. Matsumoto, H. Nagaoka, and F. G. S. L. Brandao, (2019), ArXiv:1907.05650.

[69] I. Bjelaković, T. Krüger, R. Siegmund-Schultze, and A. Szkoła, The Shannon-McMillan theorem for ergodic quantum lattice systems, Inventiones Math. **155**, 203 (2003).

[70] S. Daftuar and M. Klimesh, Mathematical structure of entanglement catalysis, Phys. Rev. A **64**, 042314 (2001).

[71] R. Landauer, Irreversibility and heat generation in the computing process, IBM J. Res. Dev. **5**, 183 (1961).

[72] C. H. Bennett, Notes on Landauer's principle, reversible computation, and Maxwell's demon, Stud. History Philosophy Sci. Part B: Stud. History Philosophy Mod. Phys. **34**, 501 (2003).

[73] S. Turgut, (2007), arXiv preprint ArXiv:0707.0444.

[74] M. Klimesh, (2007), arXiv preprint ArXiv:0709.3680.

[75] D. Jonathan and M. B. Plenio, Entanglement-Assisted Local Manipulation of Pure Quantum States, Phys. Rev. Lett. **83**, 3566 (1999).

[76] M. P. Müller, Correlating Thermal Machines and the Second Law at the Nanoscale, Phys. Rev. X **8**, 041051 (2018).

[77] P. Boes, R. Gallego, N. H. Y. Ng, J. Eisert, and H. Wilming, By-passing fluctuation theorems, Quantum **4**, 231 (2020).

[78] P. Boes, J. Eisert, R. Gallego, M. P. Müller, and H. Wilming, Von Neumann Entropy from Unitarity, Phys. Rev. Lett. **122**, 210402 (2019).

[79] H. Wilming, "Entropy and reversible catalysis," (2020), ArXiv:2012.05573.

[80] In fact, for any state $\hat{\rho}$ such that $D(\rho, \hat{\rho}) \leq \epsilon$, we have that $\hat{\rho} \succ_\sigma \rho_{\text{fl}}^\epsilon$, and therefore $\rho_{\text{fl}}^\epsilon$ is also known as the flattest state. The analogous statement is however not true for $\rho_{\text{st}}^\epsilon$, as Ref. [24] shows that there is no unique ϵ-steepest state in general.

[81] N. Datta, Min- and max-relative entropies and a new entanglement monotone, IEEE Trans. Inf. Theory **55**, 2816 (2009).

[82] K.-S. Song, Rényi information, loglikelihood and an intrinsic distribution measure, J. Stat. Plan. Inference **93**, 51 (2001).

[83] A. Winter, Tight uniform continuity bounds for quantum entropies: Conditional entropy, relative entropy distance and energy constraints, Commun. Math. Phys. **347**, 291 (2016).

[84] D. Markham, J. A. Miszczak, Z. Puchała, and K. Życzkowski, Quantum state discrimination: A geometric approach, Phys. Rev. A **77**, 042111 (2008).

[85] A. Capel, A. Lucia, and D. Pérez-García, Superadditivity of quantum relative entropy for general states, IEEE Trans. Inf. Theory **64**, 4758 (2017).

[86] H. Wilming, R. Gallego, and J. Eisert, Axiomatic characterization of the quantum relative entropy and free energy, Entropy **19**, 241 (2017).

[87] T. van Erven and P. Harremos, Rényi divergence and Kullback-Leibler divergence, IEEE Trans. Inf. Theory **60**, 3797 (2014).

[88] S. Flammia, "When are probability distributions completely determined by their moments?" MathOverflow, https://mathoverflow.net/q/4787 (version: 2017-04-13).

[89] Wikipedia contributors, "Newton's identities — Wikipedia, the free encyclopedia," (2020), https://en.wikipedia.org/w/index.php?title=Newton%27s`identities&oldid=962139559 [Online; accessed 27-August-2020].