arXiv:2009.08391v2 [quant-ph] 19 Oct 2020

The variance of relative surprisal as single-shot quantifier

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

[1]*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*
[2]*Institute for Theoretical Physics, ETH Zurich, 8093 Zurich, Switzerland*

(Dated: October 20, 2020)

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 i. i. d. 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 single-shot state-transitions in generic resource theories 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 sub-additivity. 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 asymptotic i. i. d. 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 information theory. 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 system is finite. Conceptually, the most extreme weakening of the i. i. d. 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 i. i. d. setting which is concerned with infinitely many independent iterations. By now there exists a detailed understanding of an intuitive trade-off between the above settings: While the i. i. d. setting, due to its various assumptions, can usually be characterized by variants of the (quantum) relative entropy,

$$S(\rho\|\sigma) := -\operatorname{tr}(\rho(\log(\rho) - \log(\sigma)),$$

the lack of structure in the single-shot setting implies that processes are more to difficult to characterize and involve a large number of independent constraints that are less convenient to work with than the well-behaved relative entropy. Here, *smoothed entropies* have turned out to be a powerful tool to describe these constraints and operationally characterize a variety of single-shot tasks [10–24].

* pboes@zedat.fu-berlin.de
† nelly.hy.ng@gmail.com
‡ henrikw@phys.ethz.ch

In this work, we develop an 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*,

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

where $\log(\frac{\rho}{\sigma}) \equiv \log\rho - \log\sigma$ is the *relative surprisal* of a quantum state ρ with respect to some operator 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 [25–41], 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 quantum states. These imply that state-transitions between states of low-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).

From the point of view of resource theories, this finding motivates the question whether there is a "cost" associated to obtaining states of low relative variance. We proceed to show that this is indeed the case: Under free operations, reductions of the relative variance necessarily imply a proportional reduction of the relative entropy. This finding is an implication of a resource-theoretic monotone that we construct and 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 a state, the better its 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.

The remainder of this paper is structured as follows: We derive our results in the general resource-theoretically motivated setting of σ-preserving channels. For the sake of clarity, in Section II we first provide an overview of our main results for the case of unital channels, corresponding to the resource theory of purity. The results are then shown in full generality and further discussed in Section 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.

II. OVERVIEW OF MAIN RESULTS

A. Setup and Notation

Let S be a quantum system represented on a d-dimensional Hilbert space with d fixed and finite and let $\mathcal{D}(S)$ denote the sets of density operators acting on S, respectively. We formulate our results from a resource theoretic point of view. Resource theories often begin with a framework of free operations, which identify operations that are easy for some experimenter to perform and are, in this sense, given "for free". In current resource theories, these free operations typically correspond to the set $\mathcal{C}_\sigma$ of all quantum channels from S to itself that *preserve a particular state* $\sigma \in \mathcal{D}(S)$(or sets of states), in the sense that $\mathcal{C}(\sigma) = \sigma$ for all $\mathcal{C} \in \mathcal{C}_\sigma$(for example, the thermal Gibbs state in thermodynamics [21, 42, 43]). For fixed σ we refer to such channels as *σ-preserving channels*. The set $\mathcal{C}_\sigma$ induces a relation $(\succ_\sigma, \mathcal{D}(S))$ on $\mathcal{D}(S)$, in which $\rho \succ_\sigma \rho'$ holds if and only if there exists a σ-preserving channel $\mathcal{C}$ such that $\mathcal{C}(\rho) = \rho'$. In this case we say that ρ "*σ-majorizes*" ρ'. This pre-order forms the backbone of many resource theories and is also what we are concerned with here.

Before we present general results for σ-preserving channels, we first provide in this section an overview of our results for the particular choice $\sigma = \mathbb{I} := \mathbb{1}_d/d$, i.e. for the set of channels $\mathcal{C}_\mathbb{I}$ that preserve the maximally mixed state on S. This choice of σ corresponds to the resource theory of purity, also known as the resource theory of stochastic non-equilibrium [43], and captures the essential insights from our results while being easier to state.

$\mathbb{I}$-preserving channels are also called *unital channels* [1] and the ordering $\succ_\mathbb{I}$, which we simply denote by $\succ$, is known as *majorization* [2]. 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. Finally, since we are often interested in approximate state transitions, we say that ρ *ϵ-majorizes* ρ', denoted by $\rho \succ_\epsilon \rho'$, if the transition can be approximated to ϵ-precision in trace distance, that is, if there exists a state ρ'_ϵ such that $\rho \succ \rho'_\epsilon$ and $D(\rho', \rho'_\epsilon) := \frac{1}{2}\|\rho' - \rho'_\epsilon\|_1 \leq \epsilon$.

B. 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)*

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

where $S(\rho) := \log(d) - S(\rho\|\mathbb{I}) = -\operatorname{tr}(\rho\log(\rho))$ 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 [44–46].

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. [47]). 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):
$$|V(\rho) - V(\rho')|^2 \leq K \log(d)^2 \cdot D(\rho, \rho')$$
for a constant K.
4. Correction to subadditivity (Lemma 11):
$$V(\rho) \leq V(\rho_1) + V(\rho_2) + K' \log(d)^2 \cdot f(I_\rho),$$
for any bipartite state ρ with respective marginal states ρ_1, ρ_2 and mutual information I_ρ, with constant K' and $f(x) = \max\{\sqrt[4]{x}, x^2\}$.

[1] 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 we are only interested in strict preservation of the input state.

[2] There exist various equivalent definitions of majorization between quantum states. At the level of d-dimensional probability distributions, we say that $\vec{p} \succ \vec{q}$ iff 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 that ρ majorizes ρ' when the vector of eigenvalues of ρ majorizes that of ρ'.

5. $V(\rho) = 0$ if and only if all non-zero 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 [48]

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

with r being the unique solution to

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

We have $\frac{1}{4} \log(d-1)^2 < V(\hat{\rho}_d) < \frac{1}{4} \log(d-1)^2 + 1/\ln(2)^2$, 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.

C. Sufficient criteria for single-shot state transitions

It is well known that when considering the i. i. d. 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

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

However, this is not the case in a single-shot setting. Here, the question whether $\rho \succ \rho'$ 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*

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

then $\rho \succ_\epsilon \rho'$.

We emphasize that this result is a fully single-shot result. It shows that the variance quantifies the single-shot deviation from the above i. i. d. 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 $\rho \succ_\epsilon \rho'$ 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. When applied to state transitions under unital channels in the i. i. d. 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 $\rho^{\otimes n} \succ_\epsilon (\rho')^{\otimes n}$ 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 i. i. d. case, where we use Result 1 (and its σ-preserving generalization) to study convertibility between sequences of n i. i. d. 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. [40, 41], namely that second-order corrections vanish when

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

with the simple proof of 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 [49, 50]. Finally, let us note that Eq. (3) is in general not very tight, since we know from Hoeffding-type bounds that, in the i. i. d. 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.

D. 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 semi-classical setting [21, 52]. 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. [10, 14]) and, in a sense, quantify complementary single-shot properties of quantum states. These quantities, the precise definition of which is given in Section 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.

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. [49] 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 [51]

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

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 used $\epsilon^2 \leq \sqrt{\epsilon}$. 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.

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

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

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 non-equilibrium states [21, 42]. 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.

E. 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)**.** *A function* $f : \mathcal{D}(S) \to \mathbb{R}$ *is called a* resource monotone *with respect to the relation* $\succ_\sigma$ *if either* $\rho \succ_\sigma \rho'$ *implies* $f(\rho) \geq f(\rho')$ *for all* ρ, ρ' *or it implies* $f(\rho) \leq f(\rho')$ *for all* ρ, ρ'*.*

Non-increasing resource monotones with respect to majorization are called *Schur convex* and non-decreasing 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 $\rho \not\succ \rho'$ but the former is often far easier to check than the latter. An examples of a Schur convex function is the purity $\mathrm{tr}(\rho^2)$ 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 $0 \leq M(\rho) \leq (\frac{1}{\ln(2)} + \log(d))^2$. Notably, unlike many commonly used monotones,

M is not additive with respect to product states. Fig. 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 $\rho \succ \rho'$, then*

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

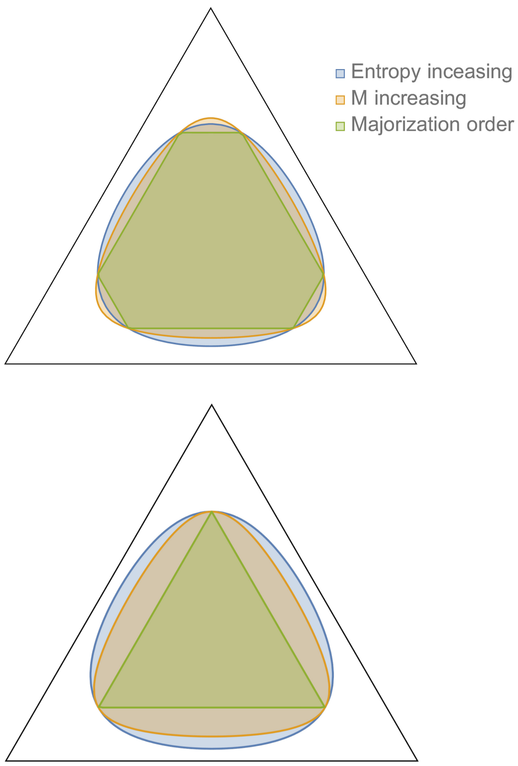

FIG. 1. Region plots of increasing M and entropy S for two initial states (top: $p = (0.65, 0.25, 0.1)$; bottom: $p = (0.7, 0.15, 0.15)$) in the simplex of states $\rho = \sum_{i=1}^{3} p_i |i\rangle\langle i|$, i.e., for $d = 3$ and fixed eigenbasis. In orange and blue are shown the sets of states with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

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 i. i. d. 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 non-trivial 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.

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 [53, 54]. 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

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

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,

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

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.

F. 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

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

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 $\mathcal{E}$ and denote $d_{SE} = d_S \cdot d_E$. Then, for any $\rho_S \in \mathcal{D}(S)$,*

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

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* [55–57]. Clearly,

$$\rho \succ \rho' \Rightarrow \rho \succ_T \rho' \Rightarrow \rho \succ_C \rho',$$

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 $S(\rho') > S(\rho)$ [58], a result that has found applications in the context of fluctuation theorems in (quantum) thermodynamics [59] and has further been strengthened in Ref. [60]. 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.

G. Local monotonicity and entropy

Result 4 is non-trivial 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 sub-additive. 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 $\rho_1' = \mathrm{tr}_2\, \mathcal{U}(\rho_1 \otimes \rho_2)$ 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

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

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 σ-PRESERVING 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, namely for σ-preserving channels. 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 that we denote as $\mathcal{C}_\sigma$ the set of all σ-preserving quantum channels from S to itself and that we say that ρ "σ-*majorizes*" ρ', writing $\rho \succ_\sigma \rho'$, whenever there exists a σ-preserving channel $\mathcal{C}$ such that $\mathcal{C}(\rho) = \rho'$. A well-known example is β-majorization in thermodynamics, where σ is the thermal state of the system at inverse temperature β [42]. As in the case of majorization, we further write $\rho \succ_{\sigma,\epsilon} \rho'$ if there exists a state ρ_ϵ' such that $D(\rho', \rho_\epsilon') \leq \epsilon$ and $\rho \succ_\sigma \rho_\epsilon'$.

Throughout, we focus on the *quasi-classical* setting, in which all states commute with σ. In particular, given σ, we will be considering elements from the set $\mathcal{S}_\sigma = \{\rho \in \mathcal{D}(S) | [\rho, \sigma] = 0\}$. Note that two states $\rho, \rho' \in \mathcal{S}_\sigma$ do not necessarily commute with one another, therefore in this sense our results cover some genuinely quantum settings (for example, in the unital case). In the following, we call a σ-preserving channel *incoherent* if $\mathcal{E}[\mathcal{S}_\sigma] \subseteq \mathcal{S}_\sigma$.

1. Lorenz curves

A key tool in proving our sufficiency results is a well-known connection between σ-majorization 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 w.l.o.g. 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* as a central tool to characterize σ-majorization.

Definition 3 (Lorenz curves)**.** *Given a quantum state σ and $\rho \in \mathcal{S}_\sigma$, the Lorenz curve $\mathcal{L}_{\rho|\sigma}(x) : [0,1] \to [0,1]$ is given by the piecewise linear curve that connects the points given by*

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

For states $\rho \notin \mathcal{S}_\sigma$ 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 equivalence relation between Lorenz curves and σ-majorization.

Theorem 4 ([61])**.** *For states* $\rho, \rho' \in \mathcal{S}_\sigma$, *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. $\rho \succ_\sigma \rho'$.

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. [62] 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 [3]

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

The states are constructed as follows.

Definition 5 (Flat approximation relative to σ)**.** *Let* σ, ρ *be density operators on a* d*-dimensional Hilbert space* $\mathcal{H}$, *where* ρ *commutes with* σ, *and* $\{|i\rangle\}_{i=1}^d$ *a common eigenbasis of the two operators that orders* ρ *relative to* σ, *yielding* $\sigma = \sum_i s_i |i\rangle\langle i|$ *and* $\rho = \sum_i^d p_i |i\rangle\langle i|$. *For any* $0 \leq \epsilon \leq 1$, *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* $M \in \{1, 2, \cdots, d-1\}$ *as the smallest integer such that*

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

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

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

[3] In fact, for any state $\hat{\rho}$ such that $D(\rho, \hat{\rho}) \leq \epsilon$, we have that $\hat{\rho} \succ_\sigma \rho_{\mathrm{fl}}^\epsilon$, and therefore $\rho_{\mathrm{fl}}^\epsilon$ is also known as the flattest state. The analogous statement is however not true for $\rho_{\mathrm{st}}^\epsilon$, as [62] shows that there is no unique ϵ-steepest state in general.

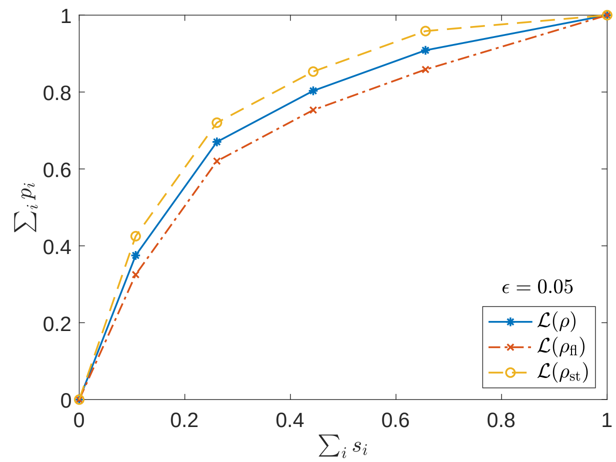

FIG. 2. An example of the Lorenz curves $\mathcal{L}(\rho)$, $\mathcal{L}(\rho_{\mathrm{st}}^\epsilon)$, and $\mathcal{L}(\rho_{\mathrm{fl}}^\epsilon)$, for some $\epsilon = 0.05$, and a state ρ of rank 5.

These integers always exist when $\epsilon \leq D(\rho,\sigma)$ *and moreover satisfy* $M \leq N$ *([62], App. D, Lemma 6). Using these definitions, finally set*

$$\bar{p}_i = \begin{cases} s_i \frac{(\sum_{j=1}^M p_j) - \epsilon}{\sum_{j=1}^M s_j}, & \text{if } i \leq M \\ s_i \frac{(\sum_{j=N}^d p_j) + \epsilon}{\sum_{j=N}^d s_j}, & \text{if } i \geq N \\ p_i & \text{otherwise.} \end{cases}$$

Definition 6 (Steep approximation relative to σ)**.** *Let* σ *be a fixed positive-definite operator and* ρ *a state that commutes with* σ, *both on a* d*-dimensional Hilbert space* $\mathcal{H}$. *Let* $\{|i\rangle\}_{i=1}^d$ *be a common eigenbasis of the two operators that orders* ρ *relative to* σ. *Then, for* $0 \leq \epsilon \leq 1$, *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* $\epsilon \leq 1 - p_1$,

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

where $R \in \{2, \ldots, d\}$ *is the largest index such that* $\sum_{i=R}^d p_i \geq \epsilon$ *and by definition of* R, *we have* $r = \sum_{i=R+1}^d p_i \leq \epsilon$. *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 to any $\rho \in \mathcal{S}_\sigma$, which is precisely what one needs to make statements regarding state transitions.

Lemma 7. *Let* ρ, σ *be two commuting quantum states acting on the same Hilbert space with* $\sigma > 0$ *and let* $1 > \epsilon > 0$.

Then, for $x \in [0,1]$,

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

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 $c = r_{\rm st}(r_{\rm fl})$ governed by both the relative entropy and its variance. These simple bounds on the Lorenz curves of $\rho_{\rm st}$ and $\rho_{\rm fl}$ are crucial for our derivation of Theorem 8 and Theorem 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 w.r.t. σ-preserving channels.

Theorem 8. *Let $\sigma > 0$ and $\rho, \rho' \in \mathcal{S}_\sigma$ be two states on S. For $1 > \epsilon > 0$, let $f_\sigma(\rho,\epsilon) := \sqrt{V(\rho\|\,\sigma)\,(2\epsilon^{-1}-1)}$. If*

$$S(\rho\|\sigma) - f_\sigma(\rho,\epsilon) > S(\rho'\|\sigma) + f_\sigma(\rho',\epsilon),$$

then $\rho \succ_{\sigma,\epsilon} \rho'$.

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

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

which implies that there exists a σ-preserving channel $\mathcal{E}$ such that $\mathcal{E}(\rho^{\bar\epsilon}_{\rm st}) = \rho'^{\bar\epsilon}_{\rm fl}$. Applying the same channel to ρ yields a state $\mathcal{E}(\rho) = \hat\rho'$ such that $D(\hat\rho',\rho') \leq \epsilon$, since

$$\begin{aligned} D(\hat\rho',\rho') &\leq D(\hat\rho',\rho'^{\bar\epsilon}_{\rm fl}) + D(\rho'^{\bar\epsilon}_{\rm fl},\rho') \\ &\leq D(\rho,\rho^{\bar\epsilon}_{\rm st}) + \bar\epsilon \leq 2\bar\epsilon = \epsilon. \end{aligned}$$

□

Result 1 in Section 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 i. i. d. state transitions with large but finite number of states, recovering a previously observed resonance condition, where second-order corrections can vanish even for non-zero 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 $\sigma > 0$ 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 $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$, then we have the following definitions [63]:

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

The *smoothed* variants are further defined as

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

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 $S^{\epsilon}_{\max}(\rho) := \log(d) - S^{\epsilon}_{\min}(\rho\|\mathbb{I})$ and $S^{\epsilon}_{\min}(\rho) := \log(d) - S^{\epsilon}_{\max}(\rho\|\mathbb{I})$ 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)**.** *Let $1 > \epsilon > 0$ and let ρ be a state on S. Then,*

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

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

Proof. We know from Ref. [62] that $S^{\epsilon}_{\max}(\rho\|\sigma) = S_{\max}(\rho^{\epsilon}_{\rm fl}\|\sigma)$. Therefore,

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

The last inequality follows from Lemma 7, which implies

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

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,

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

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

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

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

$$\mathrm{tr}(\pi_{\rho^{\epsilon}_{\rm st}}\sigma) \leq r^{-1}_{\rm st}.$$

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

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

Finally, combining Eqns. (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 $0 < \sigma \leq \mathbb{I}$ be an operator with smallest eigenvalue $s_{\min}$ on a d-dimensional Hilbert space with $d \geq 2$. For any $\rho, \rho' \in \mathcal{S}_\sigma$, if $D(\rho,\rho') \leq \epsilon \leq 1$, then*

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

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

This lemma is proven in Appendix C.

Lemma 11 (Correction to sub-additivity of relative variance.)**.** *Let $\sigma = \sigma_1 \otimes \sigma_2$ be a product state with smallest eigenvalue $s_{\min}$, on a d-dimensional, bipartite system with $d \geq 2$. Then, for any $\rho \in \mathcal{S}_\sigma$,*

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

where $K' = \sqrt{2\ln(2)}\left(12 + \log(s_{\min})^2 + 8\log(d)^2\right)$ and $f(x) = \max\{\sqrt[4]{x}, x^2\}$.

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 for fixed σ, the relative entropy $S(\rho\|\sigma)$ relative to σ is a non-increasing resource monotone with respect to σ-majorization, generalizing the Schur concavity of the von Neumann entropy. We then have the following generalization of Lemma 2 to the case of σ-preserving channels, which we prove in App. G.

Theorem 12. *Let $\sigma \in \mathcal{D}(S)$ be a full-rank state on S and let $\rho, \rho' \in \mathcal{S}_\sigma$. Then, $\rho \succ_\sigma \rho'$ implies $M(\rho'\|\sigma) \geq M(\rho\|\sigma)$, where*

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

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

Result 2 follows by setting $\sigma = \mathbb{I}$. Since $\sigma \in \mathcal{S}_\sigma$ clearly is the minimum of the order $\succ_\sigma$ over the set $\mathcal{S}_\sigma$, monotonicity implies that $0 \leq M(\rho\|\sigma) \leq M(\sigma\|\sigma) = (\frac{1}{\ln(2)} - \log(s_{\min}))^2$ for any state $\rho \in \mathcal{S}_\sigma$.

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. *For fixed full-rank state σ, let $\rho, \rho' \in \mathcal{S}_\sigma$ If $\rho \succ_\sigma \rho'$, then*

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

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 $M(\rho'\|\sigma) \geq M(\rho\|\,\sigma)$. Let us write $a = 1/\ln(2) - \log(s_{\min})$, so that $M(\rho\|\sigma) = V(\rho\|\sigma) + (a - S(\rho\|\sigma))^2$. Then reshuffling terms yields

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

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

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

Here, we have used 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 $\frac{1}{\beta}S(\rho\|\tau_\beta)$ denotes the non-equilibrium 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 $\sigma \equiv \sigma_S \otimes \sigma_E \in \mathcal{D}(S) \otimes \mathcal{D}(E)$ be a fixed product state. Further, let $\mathcal{E} : \mathcal{D}(S \otimes E) \to \mathcal{D}(S \otimes E)$ be a quantum channel from the joint system $S \otimes E$ (i.e. the tensor product of their respective Hilbert spaces) to itself that is defined via

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

for some σ-preserving channel $\mathcal{C}$ and initial state ρ_E of the environment. As in the previous section, for some initial state ρ_S on S, we denote as $\rho'_S = \mathcal{E}(\rho_S)$ 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, as defined in Eq. (7). We then have the following:

Theorem 14. *For fixed $\sigma = \sigma_S \otimes \sigma_E$, let $\mathcal{E}$ be defined via Eq.* (12) *with $\rho_E \in \mathcal{S}_{\sigma_E}$. Then for any "quasi-classical" state transition with $\rho_S, \rho'_S \in \mathcal{S}_{\sigma_S}$, we have*

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

where

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

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

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

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

$$\begin{aligned} S(\rho'_{SE} \| \, \sigma) &= S(\rho'_S \| \, \sigma_S) + S(\rho'_E \| \sigma_E) + I_{\rho_{SE}} \\ &\geq S(\rho'_S \| \, \sigma_S) + S(\rho'_E \| \sigma_E). \end{aligned}$$

□

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 $\sigma_1 \otimes \sigma_1$ 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*

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

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

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

with $S(\rho) = \lim_{\alpha\to 1} S_\alpha(\rho)$, $S_{\min}(\rho) = \lim_{\alpha\to\infty} S_\alpha(\rho)$ 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 i. i. d. 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 neighbourhood 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

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

where $\kappa^{(n)}$ is the n-th *cumulant of surprisal*. In Appendix I, we present the definition of the cumulants of surprisal as well as the derivation of (13) (also see [22]).

Eq. (13) is interesting for a number of reasons. To begin with, we have $\kappa^{(1)} = S(\rho)$ and $\kappa^{(2)} = V(\rho)$. Hence, (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 [22, 64]. It lets us apply some of our results to the Rényi curve. For instance, Result 2 relates the neighbourhood 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 sketch a proof of these statements 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 non-commuting matrices. We leave these questions to future work.

Acknowledgements. The authors would like to thank Angela Capel, Xavier Coiteux-Roy, Iman Marvian, Renato Renner, Carlo Sparaciari, Marco Tomamichel and Stefan Wolf for stimulating discussions and suggestions and especially Jens Eisert for fruitful comments on an earlier version of this work. P. B. and N. N. acknowledge support by DFG grant FOR 2724 and FQXi. P. B. further acknowledges funding from the Templeton Foundation and N. N. by the Alexander von Humboldt foundation. H. W. acknowledges contributions from the Swiss National Science Foundation via the NCCR QSIT as well as project No. 200020_165843.

---

[1] B. Schumacher, Physical Review A **51**, 2738 (1995).
[2] R. Jozsa and B. Schumacher, Journal of Modern Optics **41**, 2343 (1994).
[3] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Physical Review A **53**, 2046 (1996).
[4] B. Schumacher and M. D. Westmoreland, Physical Review A **56**, 131 (1997).
[5] A. Holevo, IEEE Transactions on Information Theory **44**, 269 (1998).
[6] S. Lloyd, Physical Review A **55**, 1613 (1997).
[7] P. M. Hayden, M. Horodecki, and B. M. Terhal, Journal of Physics A: Mathematical and General **34**, 6891 (2001).
[8] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, Phys. Rev. Lett. **111**, 250404 (2013).
[9] M. Horodecki, P. Horodecki, and J. Oppenheim, Phys. Rev. A **67**, 062104 (2003).
[10] R. Renner and S. Wolf, in *International Symposium on Information Theory ISIT* (IEEE, 2004).
[11] R. Renner, *Security of Quantum Key Distribution*, Ph.D. thesis, ETH Zurich (2005), arXiv:quant-ph/0512258.
[12] R. Renner and S. Wolf, in *Lecture Notes in Computer Science* (Springer Berlin Heidelberg, 2005) pp. 199–216.
[13] R. Renner, S. Wolf, and J. Wullschleger, in *2006 IEEE International Symposium on Information Theory* (IEEE, 2006).
[14] R. Konig, R. Renner, and C. Schaffner, IEEE Transactions on Information Theory **55**, 4337 (2009).
[15] F. Buscemi and N. Datta, IEEE Transactions on Information Theory **56**, 1447 (2010).
[16] M. Tomamichel, R. Colbeck, and R. Renner, IEEE Transactions on Information Theory **56**, 4674 (2010).
[17] F. G. S. L. Brandao and N. Datta, IEEE Transactions on Information Theory **57**, 1754 (2011).
[18] L. Wang and R. Renner, Physical Review Letters **108**, 200501 (2013).
[19] N. Datta, J. M. Renes, R. Renner, and M. M. Wilde, IEEE Transactions on Information Theory **59**, 8057 (2013).
[20] N. Datta, M. Mosonyi, M.-H. Hsieh, and F. G. S. L. Brandao, IEEE Transactions on Information Theory **59**, 8014 (2013).
[21] F. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, Proceedings of the National Academy of Sciences **112**, 3275 (2015).
[22] M. Tomamichel, *Quantum Information Processing with Finite Resources* (Springer International Publishing, 2016).
[23] A. Anshu, V. K. Devabathini, and R. Jain, Physical Review Letters **119**, 120506 (2017).
[24] G. Gour, Physical Review A **95**, 062314 (2017).
[25] V. Strassen, Transactions of the Third Prague Conference on Information Theory etc, 1962. Czechoslovak Academy of Sciences, Prague , 689 (1962).
[26] M. Hayashi, IEEE Transactions on Information Theory **54**, 4619 (2008).
[27] M. Hayashi, IEEE Transactions on Information Theory **55**, 4947 (2009).
[28] Y. Polyanskiy, H. V. Poor, and S. Verdu, IEEE Transactions on Information Theory **56**, 2307 (2010).
[29] M. Tomamichel and M. Hayashi, IEEE Transactions on Information Theory **59**, 7693 (2013).
[30] S. Verdu and I. Kontoyiannis, in *2012 46th Annual Conference on Information Sciences and Systems (CISS)* (IEEE, 2012).
[31] Y. Altug and A. B. Wagner, IEEE Transactions on Information Theory **60**, 4417 (2014).
[32] K. Li, The Annals of Statistics **42**, 171 (2014), arXiv:1208.1400v3.
[33] N. Datta and F. Leditzky, IEEE Transactions on Information Theory **61**, 582 (2014).
[34] V. Y. F. Tan, Foundations and Trends in Communications and Information Theory, **11** (2014), arXiv:1504.02608v1.
[35] M. Tomamichel and V. Y. F. Tan, Communications in Mathematical Physics **338**, 103 (2015).
[36] M. Tomamichel, M. Berta, and J. M. Renes, Nature Communications **7**, 11419 (2016).
[37] W. Kumagai and M. Hayashi, IEEE Transactions on Information Theory **63**, 1829 (2016).
[38] C. T. Chubb, M. Tomamichel, and K. Korzekwa, Quantum **2**, 108 (2018).
[39] C. T. Chubb, V. Y. F. Tan, and M. Tomamichel, Communications in Mathematical Physics **355**, 1283 (2017).
[40] C. T. Chubb, M. Tomamichel, and K. Korzekwa, Physical Review A **99**, (2019).
[41] K. Korzekwa, C. T. Chubb, and M. Tomamichel, Physical Review Letters **122**, 110403 (2019).
[42] M. Horodecki and J. Oppenheim, Nature communications **4**, 1 (2013).
[43] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Yunger Halpern, Phys. Rep. **583**, 1 (2015), arXiv:1309.6586.
[44] H. Yao and X.-L. Qi, Physical Review Letters **105**, 080501

(2010).
[45] J. Schliemann, Physical Review B **83**, 115322 (2011).
[46] J. de Boer, J. Järvelä, and E. Keski-Vakkuri, Physical Review D **99**, 066012 (2019).
[47] F. Dupuis and O. Fawzi, IEEE Transactions on Information Theory **65**, 7596 (2019).
[48] D. Reeb and M. M. Wolf, IEEE Trans. Inf. Theory **61**, 1458 (2015).
[49] P. Faist, T. Sagawa, K. Kato, H. Nagaoka, and F. G. Brandão, Physical Review Letters **123**, 250601 (2019).
[50] T. Sagawa, P. Faist, K. Kato, K. Matsumoto, H. Nagaoka, and F. G. S. L. Brandao, (2019), arXiv:1907.05650.
[51] I. Bjelaković, T. Krüger, R. Siegmund-Schultze, and A. Szkoła, Inventiones mathematicae **155**, 203 (2003).
[52] S. Daftuar and M. Klimesh, Phys. Rev. A **64**, 042314 (2001).
[53] R. Landauer, IBM journal of research and development **5**, 183 (1961).
[54] C. H. Bennett, Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics **34**, 501 (2003).
[55] S. Turgut, arXiv preprint arXiv:0707.0444 (2007).
[56] M. Klimesh, arXiv preprint arXiv:0709.3680 (2007).
[57] D. Jonathan and M. B. Plenio, Phys. Rev. Lett. **83**, 3566 (1999).
[58] M. P. Müller, Phys. Rev. X **8**, 041051 (2018).
[59] P. Boes, R. Gallego, N. H. Y. Ng, J. Eisert, and H. Wilming, Quantum **4**, 231 (2020).
[60] P. Boes, J. Eisert, R. Gallego, M. P. Müller, and H. Wilming, Physical Review Letters **122**, 210402 (2019).
[61] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities : theory of majorization and its applications* (Springer Science+Business Media, LLC, 2011).
[62] R. van der Meer, N. H. Y. Ng, and S. Wehner, Phys. Rev. A **96**, 062135 (2017).
[63] N. Datta, IEEE Transactions on Information Theory **55**, 2816 (2009).
[64] K.-S. Song, Journal of Statistical Planning and Inference **93**, 51 (2001).
[65] A. Winter, Commun. Math. Phys. **347**, 291 (2016).
[66] D. Markham, J. A. Miszczak, Z. Puchała, and K. Życzkowski, Physical Review A **77** (2008), 10.1103/physreva.77.042111.
[67] H. Wilming, R. Gallego, and J. Eisert, Entropy **19**, 241 (2017).
[68] S. Flammia, "When are probability distributions completely determined by their moments?" MathOverflow, `https://mathoverflow.net/q/4787` (version: 2017-04-13).
[69] 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].

Appendix A: Overview

We begin in Appendix B by establishing all the notation used throughout our proofs and collecting a handful of technical lemmas. In Appendices C and D, we then present the proofs for Lemma 10 and 11 respectively. Next, 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 i. i. d. 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 will make frequent use of the following definitions:

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

We also remind the reader that we use logarithms with base 2, $\log = \log_2$. A few technical lemmas are 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* iff $\rho = \sigma$.

We will also use the following generalization of the Fannes-Audenaert inequality, which is implied by Lemma 7 in [65]:

Lemma 18 (Continuity of relative entropy)**.** *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, $S(\rho\|\sigma) \geq \frac{1}{2\ln(2)}\left\|\rho - \sigma\right\|_1^2$*

The following Lemma will be useful, as it will allow us to concentrate on commuting density matrices. It shows the following statement: If a density matrix that commutes with σ is mapped to ρ' by a σ-preserving channel, then there always exists a unitary commuting with σ so that i) $U\rho'U^\dagger$ commutes with the initial state ρ and ii) the trace-distance between ρ and $U\rho'U^\dagger$ is at most given by the trace-distance between ρ and ρ'.

Lemma 20. *For a positive-definite operator σ on a d-dimensional Hilbert space, let $\mathcal{E}$ be a σ-preserving channel and $\rho \in \mathcal{S}_\sigma$. Then if $\mathcal{E}(\rho) \in \mathcal{S}_\sigma$, there exists a σ-preserving unitary channel $\mathcal{U}$ such that 1) $[\rho, \mathcal{U} \circ \mathcal{E}(\rho)] = 0$, and 2) $D(\rho, \mathcal{U} \circ \mathcal{E}(\rho)) \leq D(\rho, \mathcal{E}(\rho))$.*

Proof. We write $\sigma = \oplus_i s_i \mathbb{1}_i$, where $\mathbb{1}_i$ is the identity operator in the i-th eigenspace of σ. Similarly, we can write $\rho = \oplus_i \rho_i$ and $\mathcal{E}(\rho) := \rho' = \oplus \rho_i'$. 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 [66] we directly find

$$\begin{aligned} D(\rho, \mathcal{U}(\rho')) &= \sum_i D(\rho_i U_i \rho_i' U_i^\dagger) = \sum_i D(\mathbf{p}_i, \mathbf{p}_i') \\ &\leq \sum_i D(\rho_i, \rho_i') = D(\rho, \rho'). \end{aligned}$$

□

Lemma 21. *Let $\mathcal{E}$ be a quantum channel and $\rho = \sum_i^d p_i|i\rangle\langle i|, \rho' = \sum_i^d q_i|i\rangle\langle i|, \sigma = \sum_i s_i|i\rangle\langle i|$ be three commuting states such that $\mathcal{E}(\rho) = \rho'$ and $\mathcal{E}(\sigma) = \sigma$. Then there exists a right stochastic $d \times d$ matrix E, that is, a matrix with all non-negative entries and each of whose rows sums up to 1, such that*

$$\begin{aligned} pE &= q, \\ sE &= s \end{aligned}$$

where $p = (p_1, \ldots, p_d)$ and $q = (q_1, \ldots, q_d)$ and $s = (s_1, \ldots, s_d)$.

Lemma 22 (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 23 (Domination for definite integrals)**.** *Let f, g be continuous functions. If $f(x) \geq g(x)$ in the interval $[a,b]$, then*

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

Lemma 24. *Let $x, y \in [0,1]$ such that $|x-y| \leq \frac{1}{2}$. Then*

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

Proof. The statement is clearly true if $x = y$. Hence, w.l.o.g. let $x > y$, and set $z = |x - y| = x - y \leq 1/2$. 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,

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

To show this, let us begin by evaluating the derivative $\eta'(x) = \frac{-1}{\ln(2)}[\ln(x)+1]$, 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 $[0, z]$.

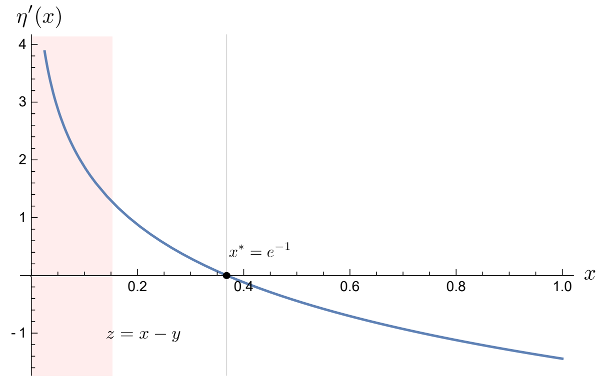

FIG. 3. The function $\eta'(x)$ and the corresponding area for the integral $F_z(0)$.

We now show that this is indeed the case. We consider three different cases: If $y \leq x^* - z$, the statement is automatically true because η' is positive and monotonically decreasing below x^*, so that an application of Lemma 23 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 applied 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 25. *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 24. 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

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

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

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

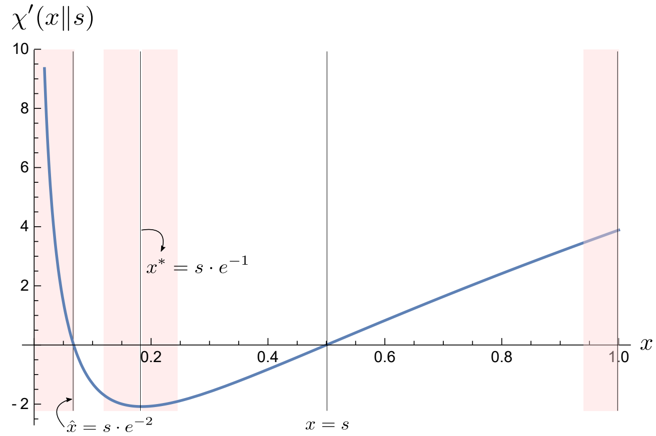

FIG. 4. The function $\chi'(x\|q)$ for $q = 0.5$, with the relevant integrals highlighted in red.

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 as wide as z) and invoking Lemma 23, 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 23, 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

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

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

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

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

Appendix C: Proof of uniform continuity of the relative variance (Lemma 10)

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

Lemma 26. *Given a bounded positive-definite operator $0 < \sigma \leq \mathbb{I}$ with smallest eigenvalue $s_{\min}$ and any two states $\rho, \rho' \in \mathcal{S}_\sigma$ on a d-dimensional Hilbert space, if $D \equiv D(\rho, \rho') \leq 1/(2e^2)$, then*

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

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

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

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 only need to consider the spectra of the various states. This follows from Lemma 20. Let $\mathcal{U}$ denote the σ-preserving unitary channel such 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

$$\begin{aligned} D(\rho,\rho') &= \frac{1}{2}\sum_i |p_i-q_i| = \frac{1}{2}\sum_i x_i, \\ |L(\rho\|\sigma)-L(\rho'\|\sigma)| &= \Big|\sum_i \chi(p_i\|s_i)-\chi(q_i\|s_i)\Big| \leq \sum_i |\chi(p_i\|s_i)-\chi(q_i\|s_i)|, \end{aligned} \tag{C1}$$

and where we have introduced the variable $x_i := |p_i-q_i|$. We now show that each of the terms in the RHS of (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 25 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

$$\begin{aligned} |\chi(p_i\|s_i)-\chi(q_i\|s_i)| &\leq \sup_{r\in[s_i/e^2,1]} |\chi'(r\|s_i)|\cdot|p_i-q_i| \\ &= \max\{|\chi'(s_i/e\|s_i)|,|\chi'(1\|s_i)|\}\cdot x_i \\ &= \frac{1}{\ln(2)^2}\cdot\max\left\{1,\ln(s_i)^2-2\ln(s_i)\right\}\cdot x_i =: \tilde{C}(s_i)x_i, \end{aligned}$$

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

Case III: $q_i<s_i/e^2<p_i$ Here, we distinguish three sub-cases. 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 sub-cases: First, assume that $q_i^*\leq p_i$. Then we can make use of Lipschitz continuity, since

$$\begin{aligned} |\chi(p_i\|s_i)-\chi(q_i\|s_i)| &= |\chi(p_i\|s_i)-\chi(q_i^*\|s_i)| \\ &\leq \tilde{C}(s_i)|p_i-q_i^*|\leq\tilde{C}(s_i)(p_i-q_i^*)\leq\tilde{C}(s_i)(p_i-q_i)=\tilde{C}(s_i)x_i. \end{aligned}$$

Note that this sub-case 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 25 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 used 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 (C1) by either $\chi(x_i\|s_i)$ or by $C(s_i)\cdot x_i$, 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 $\ln(q)^2-2\ln(q)\leq 4\ln(q)^2$. This implies that we have the upper bound

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

Let A denote the set of indices i that we have bounded by $\chi(x_i\|s_i)$ and $B=[d]\backslash A$ those that we have bounded 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

$$T_1 := \sum_{i\in A} |\chi(p_i\|s_i)-\chi(q_i\|s_i)|, \qquad T_2 := \sum_{i\in B} |\chi(p_i\|s_i)-\chi(q_i\|s_i)|$$

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

$$\begin{aligned}\chi(x_i\|s_i) &= x_i\log(x_i)^2 + 2\eta(x_i)\log(s_i) + x_i\log(s_i)^2\\ &\le x_i\log(x_i)^2 + x_i\log(s_i)^2\\ &\le x_i\log(x_i)^2 + x_i\log(x_ie^2)^2\\ &= 2x_i\log(x_i)^2 - 4\eta(x_i) + 4x_i \le 2x_i\log(x_i)^2 + 4x_i \end{aligned} \tag{C2}$$

where in the second step we used 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

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

Plugging this into the RHS of (C2) yields

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

where

$$F_1 := \sum_{i\in A}\frac{x_i}{\Delta_1}\log\left(\frac{x_i}{\Delta_1}\right)^2 = \sum_{i\in A}\chi\left(x_i/\Delta_{\le}\|1\right), \qquad F_2 := \log(\Delta_1)\cdot\sum_{i\in A}\eta(x_i).$$

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,

$$F_1 = L(\varrho\|\mathbb{I}) = V(\varrho\|\mathbb{I}) + S(\varrho\|\mathbb{I})^2 \le 4\log(|A|)^2 \le 4\log(d)^2,$$

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

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

and that $S(\varrho\|\mathbb{I})^2 = S(\varrho)^2 \le \log(d)^2$ 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

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

which yields

$$F_2 = \sum_{i\in A}\log(\Delta_1)\cdot[\Delta_1\cdot\eta(x_i/\Delta_1) - x_i\log\Delta_1] \ge \min\{0,\Delta_1\log(\Delta_1)\log(d)\} - \Delta_1\log(\Delta_1)^2.$$

The minimization arises because we distinguish two cases: If $\Delta_1\ge 1$, then the terms $\log(\Delta_1)\Delta_1\cdot\eta(x_i/\Delta_1)$ 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 (C3) then yields

$$\begin{aligned}T_1 &\le 4\Delta_1\cdot(2\log(d)^2+4) + \chi(\Delta_1\|1) + \max\{0, 2\eta(\Delta_1)\log(d)\}\\ &\le 8D\cdot(2\log(d)^2+4) + \chi(\Delta_1\|1) + \max\{0, 2\eta(\Delta_1)\log(d)\},\end{aligned}$$

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

$$\begin{aligned}|L(\rho\|\sigma) - L(\rho'\|\sigma)| &\leq 2[C(s_{\min}) + 8\log(d)^2 + 8]\cdot D + \chi(\Delta_1\|1) + \max\{0, 2\eta(\Delta_1)\log(d)\}\\ &\leq 8(3 + \frac{1}{4}\log(s_{\min})^2 + 2\log(d)^2)\cdot D + \chi(\Delta_1\|1) + \max\{0, 2\eta(\Delta_1)\log(d)\},\end{aligned}$$

where we used

$$C(s_{\min}) + 8\log(d)^2 + 8 = \max\{4, \log(s_{\min})^2\} + 8\log(d)^2 + 8 \leq 12 + \log(s_{\min})^2 + 8\log(d)^2 =: c_1.$$

Now, if $D \leq 1/(2e^2)$, then we can use the monotonicity of $\chi(\cdot\|1)$ and η over the interval $[0, 1/e^2]$ to bound $\chi(\Delta_1\|1) \leq \chi(2D\|1)$ and $\eta(\Delta_1) \leq \eta(2D)$. 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 $x \in [0, 2]$,

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

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

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

where we defined $c_2 := (6 + 2\log(d))$. □

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.

Proof. We have

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

where we used $\max\{S(\rho\|\sigma), S(\rho'\|\sigma)\} \leq -\log(s_{\min})$ in the third step, Lemma 26 in the fourth step, Lemma 18 in the fifth step, Lemma 16 in the sixth step and $\epsilon \leq \sqrt{\epsilon}$ 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 $V \equiv V(\rho\|\sigma)$, $V' \equiv V(\rho'\|\sigma)$ and $V_1 \equiv V(\rho_1\|\sigma_1)$ and similarly for the other subsystem and other quantities, L and S. We also write $I \equiv I(1:2)_\rho$, $D \equiv D(\rho, \rho_1 \otimes \rho_2), S_\otimes = S(\rho_1 \otimes \rho_2\|\sigma) = S_1 + S_2$ and $L_\otimes = L(\rho_1 \otimes \rho_2\|\sigma) = L_1 + L_2 + 2S_1S_2$. The proof of Lemma 11 is then as follows.

Proof. We have

$$\begin{aligned}V = L - S^2 &= L - L_1 - L_2 + L_1 + L_2 - (S_1 + S_2 - I)^2\\ &= V_1 + V_2 + L - L_\otimes - 2S_1S_2 + I^2 + 2IS_\otimes\\ &= L - L_1 - L_2 + V_1 + V_2 + I^2 + 2IS_\otimes - 2S_1S_2\\ &= L - L_\otimes + V_1 + V_2 + I^2 + 2IS_\otimes\\ &\leq V_1 + V_2 + \zeta(I, D, d),\end{aligned}$$

where we used Klein's inequality and the fact that both arguments are density operators in the third step and where $\zeta(I, D, d) := L - L_\otimes + I^2 - 2\log(s_{\min})I$. Our goal is to bound this function from above in terms of the mutual information I. To do so, we can apply Lemma 26, since $[\rho, \sigma] = 0$ implies that $[\rho_1 \otimes \rho_2, \sigma] = 0$. Applying this lemma yields

$$\begin{aligned}\zeta(I, d) &\leq c_1 D + c_2\sqrt{D} + I^2 + 2\log(d)I\\ &\leq c_1'\sqrt{I} + c_2'\sqrt[4]{I} + I^2 - 2\log(s_{\min})I,\end{aligned}$$

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 26, and where in the second step we used Pinsker's inequality (Lemma 19) and the fact that $I = S(\rho\|\rho_1\otimes\rho_2)$. Finally, by noting that, for $d\geq 2$, $c_1' = \max\{c_1', c_2', 1, -2\log(s_{\min})\}$, and $f(I) = \max\{I^2, I, \sqrt{I}, \sqrt[4]{I}\}$, we obtain the bound

$$\zeta(I,D,d) \leq c_1' f(I).$$

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 27.

Lemma 27. *Given any states ρ,σ such that $[\rho,\sigma] = 0$ and $\mathrm{supp}(\rho)\subseteq\mathrm{supp}(\sigma)$, 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 28. *The functions $l_c(x) = \min(c\cdot x, 1)$ 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 29 is required to then prove Lemmas 30 and 31, which jointly give rise to Lemma 7.

Lemma 29. *Let ρ,σ be two commuting d-dimensional states that satisfy $\mathrm{supp}(\rho)\subseteq\mathrm{supp}(\sigma)$, and denote the ϵ-steep and -flat approximation of ρ w.r.t. σ as $\rho_{\mathrm{st}}^{\epsilon}$ and $\rho_{\mathrm{fl}}^{\epsilon}$, according to Definition 6 and 5, respectively. Then, for $\epsilon_2\geq\epsilon_1\geq 0$,*

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

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 [62] that the ϵ-flat construction in Definition 5 is the unique state within an ϵ-ball of states around ρ, where for any state $\rho'\in\mathcal{B}^{\epsilon}(\rho)$,

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

Since $\epsilon_2\geq\epsilon_1$ implies that $\rho_{\mathrm{fl}}^{\epsilon_1}\in\mathcal{B}^{\epsilon_2}(\rho)$, the fourth inequality holds. Lastly, consider the first ineuqality. First, note that $\rho_{\mathrm{st}}^{\epsilon_1}$ and $\rho_{\mathrm{st}}^{\epsilon_2}$ always share the same basis, so we may write them as $\rho_{\mathrm{st}}^{\epsilon_1} = \sum_{i=1}^d \hat{p}_i^{(1)}|i\rangle\langle i|$ and $\rho_{\mathrm{st}}^{\epsilon_2} = \sum_{i=1}^d \hat{p}_i^{(2)}|i\rangle\langle i|$ respectively. Furthermore, they have the same relative ordering w.r.t. σ. In other words, the discrete points that define the respective Lorenz curves are aligned w.r.t. 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 $k\in\{1,\cdots,d\}$:

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

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:

$$\begin{aligned}
k = 1 \qquad & \hat{p}_1^{(2)} = p_1 + \epsilon_2 \geq p_1 + \epsilon_1 = \hat{p}_1^{(1)},\\
1 < k \leq R_2 \qquad & \sum_{i=1}^k \hat{p}_i^{(2)} = \epsilon_2 + \sum_{i=1}^k p_i \geq \epsilon_1 + \sum_{i=1}^k p_i = \sum_{i=1}^k \hat{p}_i^{(1)},\\
k > R_2 \qquad & \sum_{i=1}^k \hat{p}_i^{(2)} = 1 \geq \sum_{i=1}^k \hat{p}_i^{(1)}.
\end{aligned}$$

This finishes the proof. □

Lemma 30. *Let $\epsilon \in [0,1]$ and let ρ, σ be two commuting d-dimensional states that satisfy $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$, and denote the ϵ-steep approximation of ρ w.r.t. σ as $\rho_{\rm st}^{\epsilon}$ according to Definition 6. Then,*

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

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

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 $i \in \{1,\dots,d-1\}$. 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, $\epsilon^- \leq \epsilon \leq \epsilon^+$. Next, let $\rho_{\rm st}^{\epsilon^-} = \sum_i \hat{p}_i |i\rangle\langle i|$ denote the ϵ^--steep approximation of ρ relative to σ. 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

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

where the first inequality follows from Lemma 29 together with the fact that $\epsilon \geq \epsilon^-$, while the second inequality follows from Lemma 27. Hence, it remains to show that $A \geq 2^{S(\rho\|\sigma) - f_\sigma(\rho,\epsilon)}$. By positivity of $f_\sigma(\rho,\epsilon)$, this is clearly true whenever $S(\rho\|\sigma) \leq \log A$. In case $S(\rho\|\sigma) > \log A$, we have

$$\begin{aligned}
\epsilon \leq \epsilon^+ = \Pr\left(\log\left(\frac{p_i}{s_i}\right) \leq \log A \right) &= \Pr\left(S(\rho\|\sigma) - \log\left(\frac{p_i}{s_i}\right) \geq S(\rho\|\sigma) - \log A \right) \\
&\leq \frac{V(\rho\|\sigma)}{V(\rho\|\sigma) + [S(\rho\|\sigma) - \log A]^2},
\end{aligned}$$

where we used Cantelli's inequality (Lemma 22) with a random variable X distributed as $\mathrm{Prob}(X = \log(p_i/s_i)) = p_i$ and $\lambda \equiv S(\rho\|\sigma) - \log A$, and that the variance of relative surprisal is invariant under a sign flip. The claim then follows by a simple re-arrangement of the terms above. ∎

Lemma 31. *Let $\epsilon \in [0,1]$ and let ρ, σ be two commuting d-dimensional states that satisfy $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$, and denote the ϵ-flat approximation of ρ w.r.t. σ as $\rho_{\rm fl}^{\epsilon}$ according to Definition 5. Then,*

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

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

Proof. The proof is similar in structure to Lemma 30, 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 $i \in \{1,\dots,d-1\}$. Given ϵ, define

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

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

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

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

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

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, Def. 5, we can see that for our choice of ϵ^-, we have

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

which implies that

$$r_{\max} := \max_{l \in [d]} \frac{\bar{p}_l}{s_l} \geq \frac{p_{\tilde{\imath}}}{s_{\tilde{\imath}}} =: B$$

Therefore, by Lemma 29, Lemma 27 and Remark 28, we know that $\mathcal{L}_{\rho_{\mathrm{fl}}^{\epsilon}|\sigma}(x) \leq \mathcal{L}_{\rho_{\mathrm{fl}}^{\epsilon^-}|\sigma}(x) \leq \ell_{r_{\max}}(x) \leq \ell_B(x)$. 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 have

$$\begin{aligned}\epsilon \leq \epsilon^+ = \Pr\left(\log\left(\frac{p_j}{s_j}\right) \geq \log B\right) &= \Pr\left(\log\left(\frac{p_j}{s_j}\right) - S(\rho\|\sigma) \geq \log B - S(\rho\|\sigma)\right)\\ &\leq \frac{V(\rho\|\sigma)}{V(\rho\|\sigma) + [\log B - S(\rho\|\sigma)]^2},\end{aligned}$$

where we used Cantelli's inequality with a random variable X distributed as $\mathrm{Prob}(X = \log(p_i/s_i)) = p_i$ and $\lambda \equiv \log(B) - S(\rho\|\sigma) > 0$, in the last step. The claim then follows by re-arranging the terms in the above inequality. $\square$

Appendix F: Finite but large i. i. d. sequences

In this section, we apply Theorem 8 to study sufficient conditions for approximate state transitions in the case of n i. i. d. systems. 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* [39].

Lemma 32. *Let ρ, ρ' and σ be density matrices satisfying $[\rho,\sigma] = [\rho',\sigma] = 0$, and that $\mathrm{supp}(\rho), \mathrm{supp}(\rho') \in \mathrm{supp}(\sigma)$. 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

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

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 have denoted $r = S/S'$ 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

$$\begin{aligned} R &> \frac{1}{4}\left[2k' + 4(r-\sqrt{k}) + 2\sqrt{k'}\sqrt{k'+4(r-\sqrt{k})}\right] \\ &> r - \sqrt{k} + \sqrt{rk'}\sqrt{1 + O(1/\sqrt{\epsilon_n n})} + O(1/\epsilon_n n). \end{aligned}$$

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 32 has been derived in [40, 41] 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 [40, 41]. We should stress, however, that Lemma 32 is far from providing optimal moderate-deviation bounds for the general state-interconversion problem [40, 41]. This is to be expected, since we use a single-shot result that is not tailored to the particular structures appearing in the i. i. d. 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 33. *Let $x, y, s \in \mathbb{R}^d$ be d-dimensional row vectors with $s > 0$. Further, 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 $F(x\|s) = \sum_i s_i g\left(\frac{x_i}{s_i}\right)$ where g is a function that is concave over the interval $[\min_i \frac{x_i}{s_i}, \max_i \frac{x_i}{s_i}]$.

Proof. This proposition is almost the same as stated in Proposition B.3 of [61], pg 586, except that in [61] it holds when the function g is concave. Nevertheless, one should note that it is sufficient for g to be concave over the interval where it is evaluated, namely over the interval containing all possible input values x_i/s_i and y_i/s_i. This is because given such a g, one can always construct a continuously differentiatble 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 smallest and largest relative eigenvalues satisfy $\min_i \frac{y_i}{s_i} \geq \min_i \frac{x_i}{s_j}$ and $\max_i \frac{y_i}{s_i} \leq \max_i \frac{x_i}{s_j}$. In other words, the interval $[\min_i \frac{y_i}{s_i}, \max_i \frac{y_i}{s_i}] \subset [\min_i \frac{x_i}{s_i}, \max_i \frac{x_i}{s_i}]$. □

We are now in a position to prove Theorem 12.

Proof. Let ρ be any element of $\mathcal{S}_\sigma$. Since $[\rho,\sigma]=0$, we can express the value $M(\rho\|\sigma)$ in terms of the eigenvalues of $\rho=\sum_i p_i|i\rangle\langle i|$ and $\sigma=\sum_i s_i|i\rangle\langle i|$, using the assumption that $[\rho,\sigma]=0$. Writing $a=1/\ln(2)-\ln(s_{\min})$, this yields

$$M(\rho\|\sigma)=V(\rho\|\sigma)-2aS(\rho\|\sigma)+S(\rho\|\sigma)^2+a^2=\operatorname{tr}\left[\rho\log\left(\frac{\rho}{\sigma}\right)^2\right]-2a\operatorname{tr}\left[\rho\log\left(\frac{\rho}{\sigma}\right)\right]+a^2$$
$$=\sum_i p_i\left[\log\left(\frac{p_i}{s_i}\right)^2-2a\log\frac{p_i}{s_i}\right]+a^2=\sum_i s_i f_a\left(\frac{p_i}{s_i}\right)+a^2=:F(p\|s)+a^2,$$

where $f_a(x):=x\cdot\left[\log(x)^2-2a\log(x)\right]$ and $p=(p_1,\dots,p_d)$ and $s=(s_1,\dots,s_d)$. Consider now any σ-preserving, incoherent channel $\mathcal{E}$ and the state $\rho'=\mathcal{E}(\rho)=\sum_i p_i'|\tilde{\imath}\rangle\langle\tilde{\imath}|$. Lemma 20 implies the existence of a channel $\mathcal{E}'=\mathcal{U}\circ\mathcal{E}$ such that $\mathcal{E}'(\rho)=p_i'|i\rangle\langle i|$ commutes with both ρ and σ (that the spectrum of $\mathcal{E}'(\rho)$ coincides with that of ρ' follows from the fact that $\mathcal{U}$ is unitary). By Lemma 21, this implies that there exists a right stochastic matrix E' such that $pE'=p'$, with $(p')^T=(p_1'\dots,p_d')$. Moreover, σ-preservation of $\mathcal{E}'$ also implies that $sE=sE'=s$. We can therefore apply Lemma 33 to find that $F(p\|s)\le F(p'\|s)$ whenever f_a is concave over the interval $[\min_i\frac{p_i}{s_i},\max_i\frac{p_i}{s_i}]$. Now, it is straightforward to check that for our choice of $a=\frac{1}{\ln(2)}-\log(s_{\min})$, f_a is concave over the interval $[0,1/s_{\min}]\supseteq\left[\min_i\frac{p_i}{s_i},\max_i\frac{p_i}{s_i}\right]$, since the second derivative $f_a''\le 0$. This establishes that

$$M(\rho\|\sigma)\le M(\mathcal{E}'(\rho)\|\sigma).$$

Finally, the statement of the lemma follows by noting that

$$M(\mathcal{E}'(\rho)\|\sigma)=M(\mathcal{U}\circ\mathcal{E}(\rho)\|\sigma)=M(\mathcal{U}\circ\mathcal{E}(\rho)\|\mathcal{U}(\sigma))=M(\rho'\|\sigma),$$

where we used the unitary invariance of $V(\rho\|\sigma)$ and $S(\rho\|\sigma)$ and the fact that $\mathcal{U}$ is σ-preserving. □

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 will 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

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

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

$$f(\rho_1,\sigma_1)+f(\rho_2,\sigma_2)\ge f(\rho_1',\sigma_1)+f(\rho_2',\sigma_2)\tag{H1}$$

for any such pairs (ρ_i,σ_i) and channels $C\in\mathcal{C}_{\sigma_1\otimes\sigma_2}$. We here use the $\ge$ 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 $\mathcal{C}_\mathcal{S}$ 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

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

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 equation (H1).

Lemma 34. *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 35. *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 $(\rho_1 \otimes \sigma_2, \sigma_1 \otimes \sigma_2)$, (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

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

But since $f'(\sigma_1,\sigma_1) = 0$, we find $f'(\rho_1 \otimes \sigma_2, \sigma_1 \otimes \sigma_2) \geq f'(\rho_1,\sigma_1)$. 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

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

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

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

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

Proof. Consider the pairs $(\rho_{12}, \sigma_1 \otimes \sigma_2), (\rho_1,\sigma_1)$ 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

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

where we used 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 super-additive 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 37. *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 34 – 36 now show that $g'(\rho) := -f'(\rho, \mathbb{I}/d)$ fulfills the conditions of Lemma 9 in [60], 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

$$\begin{aligned} f(\rho) &= \log(d) - (f(\rho, \mathbb{I}/d)' + f(\mathbb{I}/d, \mathbb{I}/d)) = \log(d) + aS(\rho) - a\log(d) - f(\mathbb{I}/d, \mathbb{I}/d) \\ &= aS(\rho) + b_d. \end{aligned}$$

□

Let us now consider the setting where $\mathcal{S} = \mathcal{F}$ is the set of density matrices of full rank.

Proof of theorem 15. By the arguments presented above, we find that $f(\rho, \sigma) = f'(\rho, \sigma) + f(\sigma, \sigma)$. Lemma 34 now generalizes to show that f has to be 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 $f(\sigma, \sigma) = b$ is a constant, since for any σ_1 and σ_2 in $\mathcal{F}$ we can find channels in $\mathcal{C}_\mathcal{F}$ such that $C_2[\sigma_1] = \sigma_2$ and $\sigma_1 = C_1[\sigma_2]$. Therefore, also f' is monotonic under quantum channels in $\mathcal{C}_\mathcal{F}$. Thus f' is now continuous, additive, super-additive and monotonic under quantum channels mapping full-rank states to full-rank states and hence fulfills the conditions of the main result of Ref. [67]. This implies

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

Hence

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

□

Appendix I: Relation between Rényi curve and cumulants of surprisal

Here, we sketch the relationship between the Rényi entropies and the cumulants of surprisal. Given a state $\rho = \sum_i p_i |i\rangle\langle i|$, we can interpret the suprisal as a random variable $X \equiv -\log(\rho)$ such that $\mathrm{Prob}(X = -\log(p_i)) = p_i$. It is then simple to see that the Rényi entropies have a simple relationship to the *cumulant-generating function* of X, which is defined as

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

(here defined with respect to base 2). This relationship is

$$K_X(t) = tS_{1-t}(\rho),$$

for $t \in [-\infty, 1]$. The n-th cumulant is defined as $\kappa^{(n)} := K_X^{(n)}(0)$, that is, as the n-th derivative of $K(t)$ evaluated at $t = 0$. From iterative application of the product rule, we can see that

$$K_X^{(n)}(t) = (-1)^{n-1} n S_{1-t}^{(n-1)}(\rho) + (-1)^n t S_{1-t}^{(n)}(\rho),$$

where $S_s^{(n)}(\rho)$ denotes the n-th derivative of the curve $\alpha \mapsto S_\alpha(\rho)$ evaluated at $\alpha = s$. Hence, we have $\lim_{x\to 0} S_{1+x}^{(n)}(\rho) = (-1)^n \frac{\kappa^{(n+1)}}{n+1}$, so that the Taylor expansion of $S_\alpha(\rho)$ around $\alpha = 1$ can be rewritten in terms of the cumulants of surprisal as

$$S_\alpha(\rho) = \lim_{x\to 0} \sum_{n=0}^{\infty} \frac{S_{1+x}^{(n)}(\rho)}{n!} (\alpha - 1 - x)^n = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}\kappa^{(n)}}{n!} (\alpha - 1)^{n-1} = \sum_{n=1}^{\infty} \frac{\kappa^{(n)}}{n!} (1 - \alpha)^{n-1},$$

where as usual we define $0! = 1$ and $0^0 = 1$. By exactly analogous reasoning, we can find a similar expression for the cumulants of relative surprisal and the Rényi divergences.

Appendix J: $S_k(\rho)$ for $k = 2, \ldots, d$ encode the spectrum of ρ

We here give a proof sketch showing that the Rényi entropies $S_k(\rho)$ for $k = 2, \ldots, d$ encode the spectrum of ρ. To our knowledge this result first appeared as a comment by Steve Flammia on the website mathoverflow [68], which we repeat here for the reader's convenience. Let the eigenvalues of ρ be given by p_j with $j = 1, \ldots, d$. Then we can express the power-sums of the p_j as

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

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 compute the elementary symmetric polynomials $e_j(p_1, \ldots, p_d)$ for $j = 0, \ldots, d$ using the Girard-Newton identities [69] as

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

Note that only the power-sums for $k = 1, \ldots, d$ are required to compute the elementary symmetric polynomials. Finally, we can express the characteristic polynomial of ρ as a sum over the d elementary symmetric polynomials [69]. Solving the characteristic polynomial then gives us the eigenvalues p_j of ρ.