

# Thermodynamic fragments for magic states in quantum computation

Nikolaos Koukoulekidis\*

Department of Physics, Imperial College London, London SW7 2AZ, UK

David Jennings

School of Physics and Astronomy, University of Leeds, Leeds, LS2 9JT, UK and

Department of Physics, Imperial College London, London SW7 2AZ, UK

(Dated: March 25, 2021)

[To be sharpened] Magic states are key ingredients in schemes to realise universal fault-tolerant quantum computation. Theories of magic states attempt to quantify this computational element via monotones and determine how these states may be efficiently transformed into useful forms. Here we introduce the concept of ‘fragments’, which generalise the concept of magic monotones and has a natural thermodynamic structure based on majorisation. From this perspective magic can be viewed as a form of free energy within each fragment and is constrained by relative majorisation relations but now on quasi-probability distributions. Notably this approach allows us to incorporate actual physical constraints, for example noise models with particular bias or temperature-dependent features, and study how these constrain general magic distillation protocols. In this context we present general temperature-dependent bounds on distillation rates that any theory of magic must respect. Significantly, this analysis also presents a thermodynamic context which cannot be analysed via traditional methods based on thermodynamic entropies, due to the presence of negativity, and raises novel questions in the context of statistical mechanics.

## I. INTRODUCTION AND BACKGROUND

1. Fault-tolerance [1–9]
2. Magic [10–17]
3. Bringing in majorisation [18–26]

### A. Introduction

Magic states are necessary for achieving universal quantum computation within fault-tolerant schemes. Identifying magic as a resource for quantum universality has led to several theories which try to provide a framework for its quantification and manipulation [CITE]. The main question that such a theory attempts to answer is:

*Given two magic states  $\rho$  and  $\rho'$  is there a free operation that can convert  $\rho$  to  $\rho'$ ?*

We are interested in all resource theories of magic  $\mathcal{R} = (\mathcal{F}, \mathcal{O})$  in which free operations cannot generate any amount of resource. Further denote by  $\mathcal{D}$  the set of states considered under the theory, that is the union of free and resource states. The structure of such theory is described by a partial order [CITE], hereinafter called a *pre-order*,  $\prec_{\mathcal{R}}$  between states. We write  $\rho' \prec_{\mathcal{R}} \rho$  iff there exists  $\mathcal{E} \in \mathcal{O}$  such that  $\mathcal{E}(\rho) = \rho'$ . Naturally, states may be incomparable under the given theory, meaning that there exists no free operation that converts one to the other. We further call  $\mathcal{R}' = (\mathcal{F}', \mathcal{O}')$  a *subtheory* of  $\mathcal{R}$  iff

$\mathcal{F}' \subseteq \mathcal{F}$  and  $\mathcal{O}' \subseteq \mathcal{O}$ . The above notation will be used for general resource theories as well.

Formally, the no resource generation condition on the theories translates into two assumptions:

I Free operations send free states into free states,  $\mathcal{E} : \mathcal{F} \mapsto \mathcal{F}$ , for all  $\mathcal{E} \in \mathcal{O}$ ;

II Resource theory  $\mathcal{R}$  is a completely free state preserving theory, in the sense that for any  $d$ -dimensional ancilla system and all free operations  $\mathcal{E}$ ,  $(\mathbb{1}_d \otimes \mathcal{E})\sigma \in \mathcal{F}$  whenever  $\sigma \in \mathcal{F}$ .

The first assumption simply states that resources cannot be generated for free and is a minimal requirement for a resource theory. An immediate consequence is that if statistical mixing is included in  $\mathcal{O}$ , then the resource theory is convex. Convex resource theories have attracted a lot of attention recently [CITE] and include the magic theories discussed in Section IB. The second assumption implies that resources cannot be generated even when ancillas are allowed [example of T state generation on Bell state by Campbell].

Monotones are often used [CITE] to address the question of state convertibility, although such approaches are usually generic.

The monotonicity condition reflects the no resource generating property of free operations, so that monotones respect the pre-order  $\prec_{\mathcal{R}}$  of the theory. A monotone of any general resource theory is a projection of the theory onto the non-negative real numbers, collapsing the pre-order of the theory to the total order defined on the real line. Our contribution is the introduction of a generalised notion of *resource projection* which maps a general resource theory onto a subtheory which in principle still retains a partial structure (as opposed to the real line). Applying this notion on existing magic theories highlights

---

\* nk2314@imperial.ac.uk

the hidden stochasticity that governs magic state conversions. We show that a magic theory can be subdivided into *fragments* [\[expand\]](#)

## B. Previous work

[\[Section needs a lot of fixing\]](#)

The stabiliser theory of magic comprises of the so-called “stabiliser” states (STAB) and operations (SO), while non-stabiliser (resource) states are called magic. The stabiliser operations can be expressed in terms of a Stinespring dilation as

$$\mathcal{E}(\rho) = \text{tr}_E[U(\rho \otimes \sigma_E)U^\dagger], \quad (1)$$

for an ancilla stabiliser state  $\sigma_E$ . The motivation of the theory stems from the fact that stabiliser operations are experimentally straightforward to implement and they can be used to detect and correct errors on the stabiliser states due to their construction [\[CITE\]](#). The Gottesman-Knill theorem however indicates that stabiliser operations need to be supplemented with magic states in order to achieve universality, justifying the term “magic”.

Generalisations of the stabiliser theory appear in the literature intending to include broader classes of operations [\[CITE\]](#). The class of stabiliser preserving operations (SPO) is defined as the set of CPTP maps that send stabiliser states into stabiliser states [\[27\]](#). An important subclass of SPO is the set of completely stabiliser preserving operations (CSPO), which intuitively cannot induce “non-stabiliserness” even when applied to only part of a quantum state, i.e. operations  $\mathcal{E}$  such that  $(\mathbb{1}_d \otimes \mathcal{E})\sigma \in \text{STAB}$  for all positive dimensions  $d$  whenever  $\sigma \in \text{STAB}$ .

Even though non-stabiliserness is a necessary resource for universality, it has been proven insufficient for magic state distillation [\[15, 28\]](#). In fact, all states with non-negative Wigner distributions have been proven to be efficiently classically simulable in [\[29\]](#), a result that serves as a generalization of the Gottesman-Knill theorem. The Wigner distribution of a state in odd prime dimensions is discussed rigorously in Section [II A](#) and arises as the unique quasi-probability representation of quantum theory that identifies non-contextuality exactly with the states that are efficiently classically simulable [\[11, 30\]](#). In this framework, the stabiliser states are the only pure states represented with non-negative distributions [\[16\]](#). However, there exist mixed states with non-negative Wigner distributions that are not mixtures of stabiliser states [\[4\]](#). Therefore, stabiliser-preserving theories have been extended to a theory that preserves state “Wigner positivity” [\[12\]](#), formally defined in Section [II A](#) for odd prime dimensions. Informally, it can be considered as the maximal theory of magic  $\mathcal{R}_{\text{max}} = (\mathcal{F}_{\text{max}}, \mathcal{O}_{\text{max}})$ , where free states have non-negative Wigner distributions and free operations completely preserve this property.

[Things we MUST emphasize:](#)

1. Perhaps a nice lead-in question: “What happens if we view stabiliser states as thermodynamic equilibrium states and magic as a form of free energy?”
2. We have found a scenario in which it is impossible to describe a thermodynamic structure using any entropic approach!
3. We can tackle more ‘physicsy’ questions like: how much magic can be distilled via available operations with some given fixed-point structure?
4. This allows a diagnostic on the kind of operations needed to do good distillation. I.e. what fixed point structure should they have?
5. We go beyond the concept of monotones and replace a monotone with a  $\sigma$ -fragment.
6. We can get both upper and lower bounds on magic distillation.

## II. PHASE SPACE REPRESENTATIONS OF QUANTUM STATES

Central to our construction is the representation of any quantum state or quantum operation on a system of dimension  $d$  in terms of quasi-probability representations on a discrete phase space [CITE the review on quasi-reps]. This construction is a discrete version of Wigner representations in quantum optics.

We consider a  $d$ -dimensional quantum system with Hilbert space  $\mathcal{H}_d$ , and let  $\{|0\rangle, |1\rangle, \dots, |d-1\rangle\}$  denote the standard computational basis, defined over  $\mathbb{Z}_d = \{0, 1, \dots, d-1\}$ . On this space, generalised Pauli matrices  $X, Z$  can be defined by their respective roles as shift and phase operators, which act on the basis states as follows,

$$X |k\rangle = |k+1\rangle \quad (2)$$

$$Z |k\rangle = \omega^k |k\rangle. \quad (3)$$

Here  $\omega := e^{2\pi i/d}$  is the  $d$ -th root of unity and addition is taken modulo  $d$ . From these we can construct a phase space  $\mathcal{P}_d = \mathbb{Z}_d \times \mathbb{Z}_d$  that provides a complete representation of the quantum system. Given a point  $\mathbf{x} := (x, p)$  we define a displacement operator,

$$D_{\mathbf{x}} := \tau^{xp} X^x Z^p, \quad (4)$$

where a phase factor  $\tau := -\omega^{1/2}$  ensures unitarity. We assume going forward that  $d$  is an odd prime, however the case  $d = 2$  can be also handled, but with some additional technical caveats [CITE]. For a composite system with composite dimension  $d = d_1 \dots d_n$  we can decompose the Hilbert space as  $\mathcal{H}_d = \mathcal{H}_{d_1} \otimes \dots \otimes \mathcal{H}_{d_n}$ , and then define displacement operators as

$$D_{\mathbf{x}} := D_{(x_1, p_1)} \otimes \dots \otimes D_{(x_n, p_n)}, \quad (5)$$

where now we have

$$\mathbf{x} := (x_1, p_1, x_2, p_2, \dots, x_n, p_n) \in \mathcal{P}_{d_1} \times \dots \times \mathcal{P}_{d_n} =: \mathcal{P}_d,$$

to denote the phase space point for the composite system. For simplicity going forward we assume  $n$  copies of a  $d$ -dimensional system  $d_1 = d_2 = \dots = d$ , and therefore, we have that  $\mathbf{x} \in \mathbb{Z}_d^{2n}$ .

These displacement operators form the Heisenberg-Weyl group [CITE] under matrix multiplication modulo phases,

$$\text{HW}_d^n := \{\tau^k D_{\mathbf{x}} : k \in \mathbb{Z}_d, \mathbf{x} \in \mathcal{P}_d^n\}. \quad (6)$$

The Clifford operations  $\mathcal{C}_d^n$  are then defined as the set of unitaries that normalise the Heisenberg-Weyl group [CITE]. We may define the pure stabiliser states as those states obtained by acting on  $|0\rangle$  with Clifford unitaries. Finally, we define STAB as the convex hull of all pure stabiliser states, namely all probabilistic mixtures of states of the form  $U|0\rangle\langle 0|U^\dagger$  where  $U$  is some Clifford unitary.

## A. Wigner representations for quantum states and quantum operations

In order to provide a complete decomposition of arbitrary quantum states and quantum operations we now define a complete basis of hermitian observables that behaves naturally under the action of the Clifford group. To every point  $\mathbf{x} \in \mathcal{P}_d$  we define the phase-point operator  $A_{\mathbf{x}}$  given by

$$A_{\mathbf{x}} := \frac{1}{d} \sum_{\mathbf{z} \in \mathcal{P}_d} \omega^{\eta(\mathbf{x}, \mathbf{z})} D_{\mathbf{z}}, \quad (7)$$

where  $\eta(\mathbf{x}, \mathbf{z})$  is the symplectic inner product between any two points  $\mathbf{x}, \mathbf{z} \in \mathcal{P}_d$ , and is given explicitly by

$$\eta(\mathbf{x}, \mathbf{z}) := \mathbf{z}^T \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix} \mathbf{x}, \quad (8)$$

where  $0, \mathbb{1}$  denote the  $n \times n$  zero and identity matrices.

The phase-point operators form an orthogonal (with respect to the Hilbert-Schmidt inner product) operator basis as shown in Appendix A. Therefore, any quantum state  $\rho \in \mathcal{B}(\mathcal{H}_d)$  can be expressed as a linear combination of them,

$$\rho = \sum_{\mathbf{z} \in \mathcal{P}_d} W_{\rho}(\mathbf{z}) A_{\mathbf{z}}, \quad (9)$$

where the coefficient vector  $W_{\rho}$  is the Wigner distribution of state  $\rho$ ,

$$W_{\rho}(\mathbf{x}) := \frac{1}{d} \text{tr}[A_{\mathbf{x}} \rho]. \quad (10)$$

For any quantum state  $\rho$ , the Wigner distribution  $W_{\rho}(\mathbf{x})$  is readily seen to be a  $d^2$ -dimensional quasi-probability distribution over  $\mathcal{P}_d$  (see Appendix A for details). More precisely,  $W_{\rho}(\mathbf{x})$  is a real-valued function on  $\mathcal{P}_d$  with the property that  $\sum_{\mathbf{x}} W_{\rho}(\mathbf{x}) = 1$ . In Fig. (1), we show Wigner distributions of different types of qutrit states.

## B. Magic theories for quantum computation

The state-injection model for quantum computation [CITE] is a key avenue to realising a full-scale, fault-tolerant quantum computer. Within the approach Clifford unitaries can be done in a robust, fault-tolerant way. However, due to the Eastin-Knill theorem [CITE], it is impossible to have a universal set of transversal gates and so while Clifford unitaries can be realised transversally one must find ways around the Eastin-Knill restriction. The state-injection approach supplements the Clifford gates with noisy quantum states, called ‘magic states’, that are outside the set of stabiliser states and, when incorporated into Clifford circuits, provides universality. The obstacle to this is that the states are invariably noisy

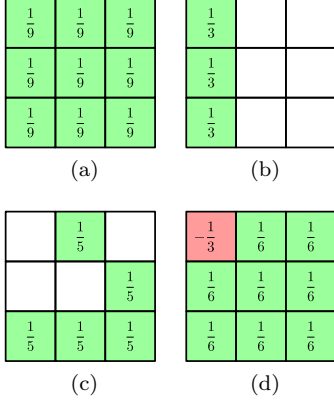


FIG. 1. **Qutrit Wigner distributions of varying magic.** (a) Maximally mixed state  $\frac{1}{3}\mathbb{1}$ ; (b) Stabilizer zero state  $|0\rangle\langle 0|$ ; (c) A non-stabiliser Wigner-positive state; (d) Magic Strange state  $|S\rangle = \frac{1}{\sqrt{2}}(|1\rangle - |2\rangle)$ , coined in [11].

and so protocols must be employed so as to purify many copies of the magic states and improve the overall performance of the induced quantum gates. A central question is then: given a number  $n$  of noisy magic states how many purified magic states can we obtain within the fault-tolerant set of quantum operations?

To address such questions both concrete distillation protocols have been developed, and there has been analysis of bounds of distillation rates within natural theoretical frameworks. These frameworks provide theories of magic in which one views magic states as “resource” states with respect to a natural class of quantum operations that are consider cheap, or “free” [CITE Gour resource review]. One very natural class of free operations considered are those obtained from Clifford operations, measurements and the ability to discard quantum systems. However there are several other candidate frameworks.

For any theory of magic a natural route to bounding the distillation rates obtainable is through the concept of a *magic monotone*. A magic monotone is a real-valued function of any quantum state  $\mathcal{M}(\rho)$  that is monotonically non-increasing under the free operations of the magic theory. More precisely  $\mathcal{M}(\sigma) \leq \mathcal{M}(\rho)$  whenever it is possible to convert  $\rho$  into  $\sigma$  using at least one of the free operations available (e.g. Clifford operations).

One of the most fundamental and commonly used magic monotones is the *mana* of a state [11], defined as

$$\text{mana}(\rho) := \ln(2\text{sn}(\rho) + 1), \quad (11)$$

where the *sum-negativity* [11] is the sum of the negative components in  $W_\rho$ ,

$$\text{sn}(\rho) := \sum_{\mathbf{x}: W_\rho(\mathbf{x}) < 0} |W_\rho(\mathbf{x})|. \quad (12)$$

Mana is an additive<sup>1</sup> magic monotone, so it provides an analytical, necessary condition for many-copy magic state interconvertibility.

A range of other monotones have been developed in the literature[CITE], and each provides a means to upper bound distillation rates within a particular theory of magic. In this work we shall consider a broad framework for magic, with conditions that any theory of magic should satisfy. However, in contrast to prior works we shall not construct magic monotones, but instead develop a generalisation to resource monotones. The approach we take will also allow us to address how distillation rates are constrained by physical properties – such as a system having biased-noise, some equilibrium structure or operations that manifest invariances, such as axial symmetry around the qubit  $Z$ -axis.

### III. STOCHASTIC STRUCTURE OF MAGIC THEORIES

We also have Wigner representations for general quantum channels. We may exploit the channel-state duality and use the normalised Choi-Jamiołkowski state

$$\mathcal{J}(\mathcal{E}) := \frac{1}{d_A}(\mathbb{1} \otimes \mathcal{E}) \sum_{i,j} |ii\rangle \langle jj| \quad (13)$$

to extend the definition of the Wigner distribution to a quantum channel  $\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \rightarrow \mathcal{B}(\mathcal{H}_B)$ , via the expression

$$W_{\mathcal{E}}(\mathbf{y}|\mathbf{x}) := d_A^2 W_{\mathcal{J}(\mathcal{E})}(\bar{\mathbf{x}} \oplus \mathbf{y}) \quad (14)$$

$$= \frac{1}{d_B} \text{tr}_B[A_{\mathbf{y}} \mathcal{E}(A_{\mathbf{x}})], \quad (15)$$

where  $\bar{\mathbf{x}} := (x, -p)$ .

The specific form of Eq. (14) is chosen so that Wigner distributions of operations act as transition matrices for Wigner distributions of states,  $W_{\mathcal{E}(\rho)} = W_{\mathcal{E}} W_\rho$ . In particular, CPTP operations that map between density operators of equal dimensions and have non-negative Wigner distributions correspond to stochastic matrices, as shown in Appendix A

#### A. Magic fragments

Equipped with the definitions of the Wigner distribution in odd prime dimensions, we can formally recast the maximal magic theory  $\mathcal{R}_{\max}$  into a stochasticity setting. The free states correspond to proper probability distributions

$$\mathcal{F}_{\max} := \{\rho : W_\rho(\mathbf{x}) \geq 0 \text{ for all } \mathbf{x} \in \mathcal{P}_d\} \quad (16)$$

<sup>1</sup> It satisfies the condition  $\text{mana}(\rho_1 \otimes \rho_2) = \text{mana}(\rho_1) + \text{mana}(\rho_2)$  which is practical in distillation settings.

The free operations should send the set of free states  $\mathcal{F}_{\max}$  into itself and completely preserve the non-negativity of the states, in the sense that  $\mathcal{E} \in \mathcal{O}_{\max}$  iff  $(\mathbb{1}_d \otimes \mathcal{E})\sigma \in \text{STAB}$  for all odd prime dimensions  $d$  whenever  $\sigma \in \mathcal{F}_{\max}$ . It is shown by Wang *et al.* [12] that  $\mathcal{O}_{\max}$  coincides with the set of operations  $\mathcal{E}$  with stochastic Wigner distributions,

$$\mathcal{O}_{\max} = \{\mathcal{E} : W_{\mathcal{E}}(\mathbf{y}|\mathbf{x}) \geq 0 \text{ for all } \mathbf{x}, \mathbf{y} \in \mathcal{P}_d\}. \quad (17)$$

Every established magic theory  $\mathcal{R} = (\mathcal{F}, \mathcal{O})$  is a sub-theory of  $\mathcal{R}_{\max}$  in the sense that  $\mathcal{F} \subseteq \mathcal{F}_{\max}$  and  $\mathcal{O} \subseteq \mathcal{O}_{\max}$ , and as such it falls under this new stochasticity setting. For technical simplicity in what follows we assume that  $\mathcal{F}$  is a closed set, and note that  $\mathcal{F}_{\max}$  is itself a closed set, since it is specified by a finite set of linear constraints of the form  $\text{tr}[L\rho] \geq 0$  with operators  $L \in \mathcal{B}(\mathcal{H})$  ensuring that the state is positive and Wigner-positive.

Given this context we now define the following key notion, that is central to our analysis.

**Definition 1 ( $\sigma$ -fragment).** *Given a resource theory of magic  $\mathcal{R} = (\mathcal{F}, \mathcal{O})$ , the  $\sigma$ -fragment of  $\mathcal{R}$  is the resource theory  $\mathcal{R}_{\sigma} = (\mathcal{F}_{\sigma}, \mathcal{O}_{\sigma})$ , where  $\mathcal{F}_{\sigma} = \{\sigma\}$  and the free operations are restricted to the ones that leave  $\sigma$  invariant,*

$$\mathcal{O}_{\sigma} := \{\mathcal{E} \in \mathcal{O} : \mathcal{E}(\sigma) = \sigma\}. \quad (18)$$

Note that  $\mathcal{F} = \bigcup_{\sigma \in \mathcal{F}} \mathcal{F}_{\sigma}$  trivially. With this basic notion defined, we now show that this union holds for free operations as well and therefore that any resource theory of magic can be faithfully subdivided into  $\sigma$ -fragments, in such a way that any problem of interconversion in the parent magic theory  $\mathcal{R}$  can be analysed across the different fragments.

**Theorem 2.** *Let  $\mathcal{R} = (\mathcal{F}, \mathcal{O})$  be a theory of magic. Every operation in  $\mathcal{O}$  leaves at least one free state invariant,*

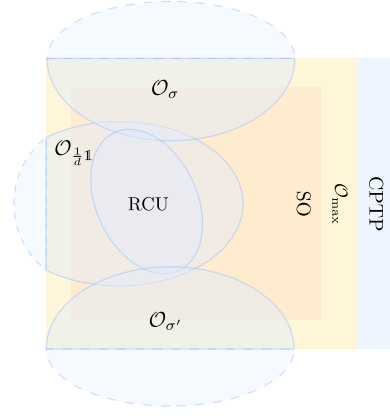
$$\mathcal{O} = \bigcup_{\sigma \in \mathcal{F}} \mathcal{O}_{\sigma}. \quad (19)$$

Therefore,  $\rho \rightarrow \tau$  in  $\mathcal{R}$  if and only if  $\rho \rightarrow \tau$  in a  $\sigma$ -fragment of  $\mathcal{R}$ .

*Proof.* Suppose  $\mathcal{E}$  is in a  $\sigma$ -fragment  $\mathcal{O}_{\sigma}$ . Then it is also in  $\mathcal{O}$ , hence  $\bigcup_{\sigma \in \mathcal{F}} \mathcal{O}_{\sigma} \subseteq \mathcal{O}$ .

Conversely, suppose  $\mathcal{E}$  is in  $\mathcal{O}$ . The free states are a closed set that is mapped one-to-one to a closed subset  $\mathcal{S}$  of the  $(d^2 - 1)$ -dimensional probability simplex.  $\mathcal{S}$  is convex, since any combination of free states is also free and the Wigner distribution is linear. Therefore,  $\mathcal{S}$  is convex and compact as a closed convex subset of the bounded compact probability simplex.

We can now view  $W_{\mathcal{E}}$  as a stochastic, continuous mapping from  $\mathcal{S}$  to itself, thus Brouwer's fixed point theorem [CITE] implies that there exists a probability distribution  $d_{\mathbf{z}}$  for some  $\mathbf{z} \in \mathcal{P}_d$  that is a fixed point of  $W_{\mathcal{E}}$ . This corresponds to a free state  $\sigma := \sum_{\mathbf{z} \in \mathcal{P}_d} d_{\mathbf{z}} A_{\mathbf{z}} \in \mathcal{F}$ . Therefore  $\mathcal{E} \in \mathcal{O}_{\sigma}$ , and so  $\mathcal{O} = \bigcup_{\sigma \in \mathcal{F}} \mathcal{O}_{\sigma}$ .



**FIG. 2. Decomposition of a magic theory  $\mathcal{R}$  into  $\sigma$ -fragments.** Examples of magic theories (SO: Stabilizer operations,  $\mathcal{O}_{\max}$ : Completely positive-Wigner-preserving operations, RCU: Random Clifford Unitaries – subclass of SO) involve operations denoted by the two yellow regions, with every other established magic theory between them. We introduce  $\sigma$ -fragments  $\mathcal{O}_{\sigma}$  defined for all free states  $\sigma$  that cover  $\mathcal{O}_{\max}$ . Each  $\mathcal{O}_{\sigma}$  is extensible to a set of stochastic maps outside the CPTP operations. Within each  $\sigma$ -fragment  $d$ -majorisation can be used allowing for an intricate, but tractable approach towards the study of magic state interconversion.

The state interconversion result follows immediately.  $\square$

The zoo of all magic operation classes is summarised in Fig. (2). Completely positive-Wigner-preserving operations [12] form the maximal operation class  $\mathcal{O}_{\max}$ . This theory of magic is thus covered by  $\sigma$ -fragments exactly and every other magic theory is contained within this cover. In particular, the stabiliser operations SO are contained within  $\mathcal{O}_{\max}$ .

Every quantum circuit aiming at a given magic state conversion  $\rho \rightarrow \tau$  possesses certain symmetries according to Theorem 2 that allow us to study the conversion within only certain  $\sigma$ -fragments that reflect these symmetries. As a simple example, the dephasing channel

$$\Delta(\rho) = \sum_{k \in \mathbb{Z}_d} |k\rangle\langle k| \rho |k\rangle\langle k| \quad (20)$$

removes coherent phases in the computational basis and therefore leaves exactly all mixtures of computational basis states invariant. A noise channel consisting only of such dephasing operations can be fully analysed in the  $\sigma$ -fragments for the pure computational basis states  $\sigma$ . [Have proven in appendix that we need to consider *only* the pure stabiliser fragments - not mixed ones]

The subdivision of magic theories into  $\sigma$ -fragments is also powerful because the pre-order  $\prec_{\mathcal{R}_{\sigma}}$  of every  $\sigma$ -fragment is described by well-behaved majorisation tools, as we establish in the rest of this section.



## B. Majorisation

Majorisation is a collection of powerful tools that has recently found many applications in quantum information theory [CITE]. It describes the [disorder / non-uniformity] of distributions that undergo stochastic transformations.

To formally state majorisation results, we first denote by  $S_d(\mathbf{d})$  the set of  $(d \times d)$  stochastic matrices that preserve the probability vector  $\mathbf{d}$ . Specifically, for any  $S \in S_d(\mathbf{d})$ , all matrix elements are non-negative, all rows sum to 1 and  $S\mathbf{d} = \mathbf{d}$ . The set  $S_d(\mathbf{d})$  forms a group under matrix multiplication for all  $\mathbf{d}$  with positive components.

Majorisation describes quantum thermodynamics exactly in the absence of quantum coherence. The use of majorisation in this setting provides vital intuition for our purposes. At any given temperature  $\beta^{-1}$ , the thermal state  $\gamma_\beta$  is thermodynamically the most disordered state. Thermal operations are defined as operations that cannot extract energy from the Gibbs state,  $\mathcal{E}(\gamma_\beta) = \gamma_\beta$ . Convertibility between states via thermal operations is equivalent to a stochasticity condition on the energy level populations of the states [CITE]. Informally, there exists a thermal operation  $\mathcal{E}$  such that  $\tau = \mathcal{E}(\rho)$  if and only if there exists a matrix  $S \in S_d(\mathbf{d})$  such that  $\mathbf{q} = S\mathbf{p}$ , where  $\mathbf{q}, \mathbf{p}$  and  $\mathbf{d}$  are the energy level population vectors of  $\tau, \rho, \gamma_\beta$  respectively. Formal statements of the above can be found [in the appendix].

Based on this setting, we define majorisation as follows.

**Definition 3 ( $\mathbf{d}$ -majorisation).** *Given  $\mathbf{x}, \mathbf{y}, \mathbf{d} \in \mathbb{R}^d$ , such that the components of  $\mathbf{d}$  are positive,  $\mathbf{y}$  is said to  $\mathbf{d}$ -majorise  $\mathbf{x}$ , iff there exists a matrix  $S \in S_d(\mathbf{d})$  such that  $\mathbf{x} = S\mathbf{y}$ .*

We denote this pre-order by  $\mathbf{x} \prec_{\mathbf{d}} \mathbf{y}$ . If  $\mathbf{d} = \frac{1}{d}\mathbf{1}$ , the  $d$ -dimensional uniform distribution, then  $S_d(\mathbf{d})$  is the set of doubly stochastic matrices and we retrieve the familiar notion of majorisation from entanglement theory. [CITE]

The pre-order imposed by  $\mathbf{d}$ -majorisation admits a numerically efficient reformulation in terms of Lorenz curves. Let the vector  $\mathbf{u}^\downarrow$  denote the vector  $\mathbf{u} \in \mathbb{R}^d$  with its components arranged in non-increasing order.

**Definition 4 (Lorenz curve).** *Let  $\mathbf{w}, \mathbf{d} \in \mathbb{R}^d$ , where the components of  $\mathbf{d}$  are positive with  $D = \sum_{i=1}^d d_i$  and denote by  $\tilde{\mathbf{w}} := (w_1/d_1, \dots, w_d/[d_d])^T$  the vector of ratios between the corresponding components of  $\mathbf{w}$  and  $\mathbf{d}$ .*

*Finally, denote by  $\pi : \mathbb{Z}_d \mapsto \mathbb{Z}_d$  the permutation that sorts  $\tilde{\mathbf{w}}$ ,  $(\tilde{\mathbf{w}}^\downarrow)_i = w_{\pi(i)}$  for all  $i = 1, \dots, d$ .*

*Consider the piecewise linear curve obtained by joining the points  $\{(0, 0)\} \cup \{(x_k, L_{\mathbf{w}|\mathbf{d}}(k))\}_{k=1, \dots, d}$ , where*

$$(x_k, L_{\mathbf{w}|\mathbf{d}}(k)) := \left( \frac{1}{D} \sum_{i=1}^k d_{\pi(i)}, \sum_{i=1}^k w_{\pi(i)} \right). \quad (21)$$

*We define the set of points on this curve,  $L_{\mathbf{w}|\mathbf{d}}(x)$ ,  $x \in [0, 1]$ , as the Lorenz curve of vector  $\mathbf{w}$  with respect to  $\mathbf{d}$ .*

Components  $x_k$  are rescaled by  $D$  so that comparison of curves with unequal dimensions is possible. In fact, the Lorenz curves  $L_{\mathbf{w}|\mathbf{d}}$  and  $L_{\mathbf{w} \otimes \mathbf{d} | \mathbf{d} \otimes \mathbf{d}}$ , where  $\otimes$  denotes the Kronecker product, are identical. Furthermore, a Lorenz curve  $L_{\mathbf{w}|\mathbf{d}}(x)$  is always concave in  $x$ , since it consists of  $d$  line segments each with slope  $(\tilde{\mathbf{w}}^\downarrow)_i$  for  $i = 1, \dots, d$  which is non-increasing by definition. Finally, the points on the interior of the Lorenz curve connecting line segments of different slopes are called *elbows*.

A vector  $\mathbf{y}$  is said to  $\mathbf{d}$ -majorise another vector  $\mathbf{x}$  if and only if the Lorenz curve  $L_{\mathbf{y}|\mathbf{d}}$  lies above Lorenz curve  $L_{\mathbf{x}|\mathbf{d}}$ , thus reducing  $\mathbf{d}$ -majorisation into a finite set of inequalities.

**Theorem 5.** *Let  $\mathbf{x}, \mathbf{y}, \mathbf{d} \in \mathbb{R}^d$ , such that the components of  $\mathbf{d}$  are positive. Then,  $\mathbf{x} \prec_{\mathbf{d}} \mathbf{y}$  if and only if  $L_{\mathbf{x}|\mathbf{d}}(x) \leq L_{\mathbf{y}|\mathbf{d}}(x)$  for all  $x \in [0, 1]$  with strict equality at  $x = 1$ .*

A restatement of the theorem including more equivalent conditions and a proof are provided in Appendix B.

## C. Majorisation of quasi-probabilities in $\sigma$ -fragments

We can approach any magic theory through a thermodynamic lens, and in doing so we are provided with valuable insights on the structure of the theory. Firstly, any free state, for example a stabiliser state, can be viewed as a thermal state  $\gamma_\beta$ . Without loss of generality we can always write a quantum state  $\sigma$  as a thermal state,  $\sigma = \gamma_\beta := \frac{1}{Z_\beta} e^{-\beta H}$  for some  $\beta \geq 0$  and Hamiltonian  $H$  (either effective or actual)<sup>2</sup>. We can also view the set of free operations  $\mathcal{O}_\sigma$  as a counterpart of thermal operations, in the sense that any operation  $\mathcal{E}$  in  $\mathcal{O}_\sigma$  preserves state  $\sigma$ .

It is then apparent that the pre-order  $\prec_{\mathcal{R}_\sigma}$  between the operations in the  $\sigma$ -fragment follows the rules of  $\mathbf{d}$ -majorisation as outlined in Section IIIB. In particular, the pre-order  $\prec_{\mathcal{R}_\sigma}$  of the  $\sigma$ -fragment  $\mathcal{R}_\sigma = (\mathcal{F}, \mathcal{O}_\sigma)$  between  $d$ -dimensional states corresponds to the majorisation pre-order  $\prec_{W_\sigma}$  between their  $d^2$ -dimensional Wigner distributions. For simplicity we shall merge the notation into  $\prec_\sigma$ , as there is little risk of confusion.

Note that the Wigner components of an  $n$ -copy state  $\rho^{\otimes n}$  can be calculated directly from  $W_\rho$  by convolution of the distribution with itself,

$$W_{\rho^{\otimes n}} = W_\rho^{\otimes n}, \quad (22)$$

where  $\otimes$  can be interpreted as the usual Kronecker product in the last expression. We may use the vector notation  $\mathbf{w}(\rho) = (w(\rho)_i)_{i=1, \dots, d^2}$  for the Wigner distribution, in which case we can express the Kronecker product as

$$W_\rho \otimes W_\rho = (w(\rho)_i w(\rho)_j)_{i,j=1, \dots, d^2}. \quad (23)$$

<sup>2</sup> Technicalities arise for the case where  $\sigma$  is not full rank, but this can be still described via  $\beta \rightarrow 0$  limiting process.

The correspondence between the phase space and pure vector representations of the Wigner distribution is discussed more in Appendix F 1. [In the rest of the paper we will be using the two notations interchangeably.]

Furthermore, we restrict our analysis to  $\sigma$ -fragments, where  $\sigma$  is full-rank. This is justified as majorisation is continuous between fragments, in the sense that the pre-orders  $\prec_\sigma$  and  $\prec_{\sigma'}$  are equivalent for states  $\sigma$  and  $\sigma'$  which are  $\epsilon$ -close. [CHECK] Therefore, majorisation analysis is robust under imperfections in the experimental implementation of quantum operations.

In particular, we can always use common noise effects to approximate any  $\sigma$ -fragment where  $\sigma$  is not full-rank by the  $\sigma'$ -fragment where  $\sigma'$  is a noisy approximation of  $\sigma$ . For example, inducing depolarising noise, we can write  $\sigma' = (1 - \epsilon)\sigma + \epsilon \frac{1}{d} \mathbb{1}$ , for some infinitesimal  $\epsilon > 0$ , so that  $\sigma'$  is full-rank and arbitrarily close to  $\sigma$ . Important examples of such  $\sigma$ -fragments include pure stabiliser states which are rank-1, e.g. the zero state depicted in Fig. (1(b)). Operations in such fragments include important stabiliser operations like the replacement channel,  $\mathcal{E}(\rho) = \sigma$  for all states  $\rho$ .

**Theorem 6.** *Let  $\mathcal{R} = (\mathcal{F}, \mathcal{O})$  be a theory of magic. If  $\rho \rightarrow \tau$  in  $\mathcal{R}$  then  $W_\tau \prec_\sigma W_\rho$  within at least one  $\sigma$ -fragment.*

*Proof.* Suppose we can convert  $\rho$  into  $\tau$  in the magic theory. Thus there is some  $\mathcal{O}_\sigma$  and some  $\mathcal{E} \in \mathcal{O}_\sigma$  such that  $\mathcal{E}(\rho) = \tau$ , and  $\mathcal{E}(\sigma) = \sigma$ . Therefore, the Wigner distribution of this free operation satisfies  $W_\mathcal{E} \in S_{d^2}(W_\sigma)$  and  $W_\mathcal{E} W_\rho = W_\tau$ . Since  $\sigma$  is full-rank and free, its Wigner distribution is strictly positive in all components, so it directly follows from Definition 3 that  $W_\tau \prec_\sigma W_\rho$ .  $\square$

The result of Theorem 6 can be understood as an extension of the idea of a magic monotone, where we replace  $\mathcal{M}(\tau) \leq \mathcal{M}(\rho)$  with  $W_\tau \prec_\sigma W_\rho$ . The physical difference between the two expressions is that the majorisation ordering occurs in a specific  $\sigma$ -fragment. Therefore, majorisation constraints can be used to place upper bounds on magic state distillation in a way that allows one to incorporate the physics of the allowed operations – specifically, it enables us to bound how much magic can be distilled via quantum operations that, for example, preserve the equilibrium state of the system, or via operations that are symmetric about the  $Z$ -axis of the Bloch sphere. We discuss distillation upper bounds in detail in Sections IV and V

This approach can also provide *lower bounds* on distillation, however now more structure about the specific free operations must be included. We briefly discuss distillation lower bounds in [REF].

[Mention relative majorisation as a possibility]

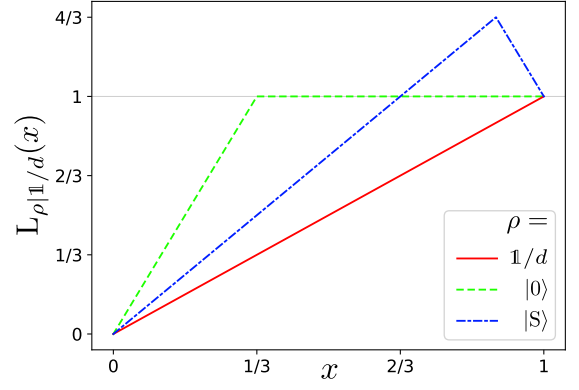


FIG. 3. **Quasi-probability Lorenz curve comparison.** The Lorenz curves are constructed by the Wigner distributions illustrated in Fig. (1) in the unital fragment  $\mathcal{O}_{\mathbb{1}/d}$  for  $d = 3$ . The maximally mixed state curve is simply the line connecting  $(0, 0)$  and  $(1, 1)$ . There is no operation in  $\mathcal{O}_{\mathbb{1}/d}$  that can convert  $|S\rangle$  to  $|0\rangle$ , as their Lorenz curves intersect.

#### D. Lorenz curves of quasi-probabilities in $\sigma$ -fragments

It is straightforward to construct Lorenz curves for Wigner distributions in any  $\sigma$ -fragment. As we have seen in Theorem 6, for any full-rank free state  $\sigma$  we have that  $W_\sigma$  is a strictly positive full-rank probability distribution, and so one can define a corresponding notion of  $d$ -majorisation on *quasi*-distributions. We write  $L_{\rho|\sigma}(x)$  for the Lorenz curve of  $W_\rho$  with respect to  $W_\sigma$ . The vector of ratios  $\tilde{w}(\rho|\sigma)$  used to construct the curve is

$$\tilde{w}(\rho|\sigma)_i := \frac{w(\rho)_i}{w(\sigma)_i}, \quad (24)$$

and is called the *rescaled* distribution of  $\rho$  with respect to  $\sigma$ .

An example of comparison between different Lorenz curves is provided in Fig. (3). The curves in the figure are constructed in the *unital fragment*, i.e. the  $\sigma$ -fragment defined by the maximally mixed state  $\sigma = \frac{1}{d} \mathbb{1}$  whose Wigner distribution is the uniform probability distribution.

Normalisation of the Wigner distribution ensures that for all quantum states  $\rho$ ,  $L_{\rho|\sigma}(x) \geq 0$  and  $L_{\rho|\sigma}(1) = 1$ . We stress that  $0 \leq L_{\rho|\sigma}(x) \leq 1$  for all  $x \in [0, 1]$  if and only if  $\rho$  is a positive Wigner state. As a consequence, checking whether a magic state conversion of the form

$$\rho \xrightarrow{\mathcal{E} \in \mathcal{O}_\sigma} \tau \quad (25)$$

is not possible, reduces to the set of constraints

$$L_{\rho|\sigma}(x) \geq L_{\tau|\sigma}(x), \quad x \in [0, 1], \quad (26)$$

due to Theorems 5 and 6.

We can refine the number of independent constraints stemming from this inequality. In fact, there are only

as many independent constraints as there are elbows in the Lorenz curve of the target state as shown in Appendix D 4. However, in principle any one location  $x$  provides a valid constraint leading to some upper distillation bound, while optimising over the location would provide the strictest bound. In particular, we establish the *first elbow constraint* which we use in Sections IV and V explicitly.

**Lemma 7.** *Consider a magic state interconversion as in Eq. (25), where we denote by  $(x_0, L_0)$  and  $(x'_0, L'_0)$  the first elbow coordinates of the initial and target states respectively. If  $x_0 < x'_0$ , then it holds that*

$$\frac{L_0}{x_0} \geq \frac{L'_0}{x'_0}. \quad (27)$$

*Proof.* Consider the Lorenz curve constraint at  $x = x_0$ ,

$$L_{\rho|\sigma}(x_0) \geq L_{\tau|\sigma}(x_0). \quad (28)$$

Since  $x_0 < x'_0$ , we can find the target state Lorenz curve coordinate  $L'_*$  at location  $x = x_0$  by interpolating between the origin and the target state's first elbow,

$$L'_* = \frac{x_0}{x'_0} L'_0. \quad (29)$$

We need  $L_0 \geq L'_*$  and rearranging we get Eq. (29).  $\square$

We can now turn the attention from the first elbow to the peak of the Lorenz curve to associate it directly with the magic monotone of mana, and equivalently the sum-negativity of a quantum state. This holds independently of the particular  $\sigma$ -fragment one works in and as a result it becomes apparent that mana provides a weaker condition than majorisation for all magic state interconversions.

We first show that the Lorenz curve maximum of state  $\rho$  is independent of the  $\sigma$ -fragment and directly related to its sum-negativity.

**Lemma 8.** *Given a quantum state  $\rho$ , the maximum of its Lorenz curve  $L_{\rho|\sigma}$  is independent of the  $\sigma$ -fragment and is equal to  $1 + \text{sn}(\rho)$ .*

*Proof.* We may use the vector notation of the Wigner distributions  $\mathbf{w}(\rho)$  and  $\mathbf{w}(\sigma)$ . We choose the component indexing so that the rescaled distribution

$$\tilde{\mathbf{w}}(\rho|\sigma) := \left( \frac{w(\rho)_1}{w(\sigma)_1}, \dots, \frac{w(\rho)_{d^2}}{w(\sigma)_{d^2}} \right)^T, \quad (30)$$

is sorted,  $\tilde{\mathbf{w}} = \tilde{\mathbf{w}}^\downarrow$ . Note that all components of  $\mathbf{w}(\sigma)$  are positive, so  $\tilde{w}_i \geq 0$  if and only if  $w(\rho)_i \geq 0$  for any  $i = 1, \dots, d^2$ .

Let  $i_*$  be the index of the smallest non-negative component of  $\tilde{\mathbf{w}}^\downarrow$ . Then,  $w(\rho)_i < 0$  if and only if  $i > i_*$ , so the maximum of Lorenz curve  $L_{\rho|\sigma}(x)$  takes the value

$$L_{\rho|\sigma}(x_{i_*}) = \sum_{i=1}^{i_*} w(\rho)_i, \quad (31)$$

and is achieved at

$$x_{i_*} := \sum_{i=1}^{i_*} w(\sigma)_i. \quad (32)$$

The location of the maximum ( $x = x_{i_*}$ ) varies from fragment to fragment, but its value is independent of  $\sigma$ ,

$$L_{\rho|\sigma}(x_{i_*}) = \sum_{\mathbf{x}: W_\rho(\mathbf{x}) \geq 0} W_\rho(\mathbf{x}) = 1 + \text{sn}(\rho). \quad (33)$$

$\square$

We can therefore view mana as just one feature of the Lorenz curve, namely its maximum value. Conversely, it is now clear that the maximum of the Lorenz curve acts as a valid magic monotone.

**Theorem 9.** *Given a magic state conversion  $\rho \rightarrow \tau$ , the majorisation condition is stronger than the mana condition in every  $\sigma$ -fragment.*

*Proof.* The maximum of the Lorenz curve of a state  $\rho$  is independent of the  $\sigma$ -fragment due to Lemma 8, and can be expressed as an increasing function of mana,

$$\max_{x \in [0,1]} L_{\rho|\sigma}(x) = 1 + \text{sn}(\rho) = \frac{1}{2} \left( 1 + e^{\text{mana}(\rho)} \right). \quad (34)$$

Therefore, the majorisation bound

$$L_{\rho|\sigma}(x) \geq L_{\tau|\sigma}(x), \quad x \in [0,1] \quad (35)$$

implies the order  $\max_{x \in [0,1]} L_{\rho|\sigma}(x) \geq \max_{x \in [0,1]} L_{\tau|\sigma}(x)$ , hence the mana condition  $\text{mana}(\rho) \geq \text{mana}(\tau)$ .  $\square$

The area  $\mathcal{A}_\sigma(\rho)$  between the curve  $L_{\rho|\sigma}$  and the line  $y = 1$  is a resource monotone in the  $\sigma$ -fragment. This is clear because for any state conversion like Eq. (25), the Lorenz curve  $L_{\tau|\sigma}$  is lower than  $L_{\rho|\sigma}$ , hence  $\mathcal{A}_\sigma(\mathcal{E}(\rho)) \leq \mathcal{A}_\sigma(\rho)$ . The exact form of the monotone is given in Appendix D 3. [should we include the area? Have a look at Appendix D 3]

In Sections IV and V, we study majorisation constraints on magic distillation arising in different fragments. For this reason, we define the  $n$ -copy,  $\epsilon$ -noisy Strange state,

$$\rho_S(\epsilon)^{\otimes n} := \left[ (1 - \epsilon) |S\rangle \langle S| + \epsilon \frac{1}{3} \mathbb{1} \right]^{\otimes n}, \quad (36)$$

in  $\mathcal{O}_\sigma$ , where the pure magic state  $|S\rangle$  is induced with depolarising noise. Its Wigner distribution is visualised in Fig. (1(d)).

We refer to parameter  $\epsilon$  as *noise level* to avoid confusion with the error rate  $\delta$ , a term commonly used in the literature to denote that the distilled state has a marginal overlap of at least  $1 - \delta$  with the desired magic state. In our case the error rate of the state in Eq. (36) would be  $\delta = \frac{2}{3}\epsilon$ .



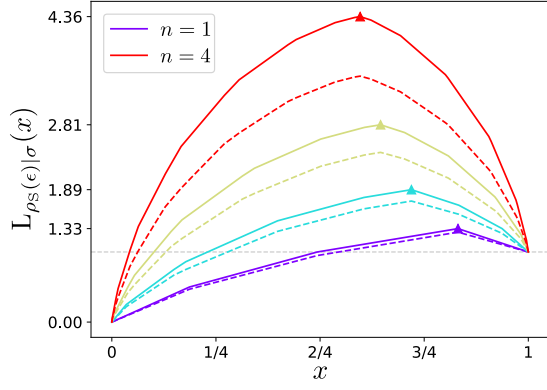


FIG. 4. **A ‘heretical’ family of Lorenz curves.** Traditionally, Lorenz curves are monotone increasing cumulant functions that reach a maximum value of 1. In contrast, Lorenz curves for magic states break through the value of  $L(x) = 1$  due to the presence of negativity in the associated quasi-probability distribution. The above family of curves correspond to multiple copies of noisy Strange states  $\rho_S(\epsilon)^{\otimes n}$  for  $n = 1, 2, 3, 4$  within a particular stabilizer fragment. Solid lines represent pure Strange states, while dashed lines represent noisy Strange states with depolarising noise  $\epsilon = 0.1$ .

At  $\epsilon = 0$ , the Strange state is a qutrit magic state of maximal sum-negativity / mana [11] and therefore acts as an ideal distillation target, analogous to a Bell state in bipartite entanglement theory. The Strange state is exceptionally symmetric under Clifford transformations and as a result there exists a *twirling* protocol, discussed in detail in [31, 32], allowing for the conversion of any noisy magic state to the form of Eq. (36) via Cliffords.

In Fig. (4), we illustrate the Lorenz curve  $L_{\rho_S|\sigma}$ <sup>3</sup> of pure and noisy  $n$ -copy Strange states in some thermal  $\sigma$ -fragment. Due to Lemma 8, it is clear that the curves peak at  $1 + \text{sn}(\rho_S(\epsilon)^{\otimes n})$ .

#### IV. MAGIC IN THE UNITAL FRAGMENT

The unital fragment encompasses the circuits which preserve the maximally mixed state ( $\mathbb{1}/d$ ) and so it includes many important families of circuits.

MSD circuits in principle consist of bulk sequences of random Clifford unitaries (RCU)[CITE], depicted in Fig. (2). Operations in RCU can be expressed as

$$\mathcal{E}(\rho) = \sum_i p_i U_i \rho U_i^\dagger, \quad U_i \in \mathcal{C}_d. \quad (37)$$

Depending on the symmetries of such operations, a Clifford sequence may belong in other  $\sigma$ -fragments as well. In such a case, the majorisation condition (26) needs to be checked in the  $\sigma$ -fragments that reflect all symmetries of the operation sequence.

In general, noisy circuits are well-described by the unital fragment. To see this, consider incorporating noisy channels in the circuit, for example dephasing channels as in Eq. (20) defined in different bases. This process destroys the circuit symmetries, except for the invariance of the maximally mixed state. Dephasing and bit-flip error channels are examples of the many error-inducing channels that respect the unital symmetry. [Expand on significance of unital fragment]

We consider the task of purifying  $n$  copies of a noisy Strange state  $\rho_S(\epsilon)^{\otimes n}$  as given in Eq. (36) into a smaller number of copies  $n'$  of a less noisy strange state  $\rho_S(\epsilon')^{\otimes n'}$ , with  $\epsilon' < \epsilon$  and  $n' \leq n$ ,

$$\rho_S(\epsilon)^{\otimes n} \longrightarrow \rho_S(\epsilon')^{\otimes n'} \otimes \left(\frac{1}{3}\mathbb{1}\right)^{\otimes(n-n')}, \quad (38)$$

where all copies  $n, n', n - n'$  are even. Since the state  $\mathbb{1}/3$  is free, tensoring in copies of it does not affect the distillation process. The distillation rate  $R := n'/n$  for this process will in general depend on the noise levels,  $R = R(\epsilon, \epsilon')$  and our task is to provide it with an upper bound.

The Lorenz curve  $L_{\rho_S|\mathbb{1}/3}$ , for some general noise parameter  $\epsilon$  and number of copies  $n$ , is defined at  $9^n$  points between 0 and 1. The exact expressions for the coordinates of these points can take 8 different forms, depending on whether the noise level  $\epsilon$  is greater or less than  $\frac{3}{7}$ , the parity of the number of copies  $n$  is even or odd and the location relative to the curve peak is on the left hand side (LHS – including the curve peak) or right hand side (RHS) of the curve peak. The full details for the construction of all Lorenz curve forms are provided in Appendix E.

Here we focus on the comparison of the first elbow (which lies in the LHS part) of Lorenz curves with even copies  $n, n'$  and low noise levels ( $\epsilon' < \epsilon \leq 3/7$ ).

The Wigner distribution of the 1-copy,  $\epsilon$ -noisy Strange state can be written as

$$W_{\rho_S(\epsilon)}(\mathbf{x}) = (1 - \epsilon)W_{|S|S|}(\mathbf{x}) + \epsilon W_{\frac{1}{3}\mathbb{1}}(\mathbf{x}), \quad (39)$$

so we get positive components

$$u(\epsilon) := \frac{1}{6} - \frac{1}{18}\epsilon \quad (40)$$

at the 8 phase space points  $\mathbf{x} \in \mathcal{P}_3 \setminus \{\mathbf{0}\}$  and a negative component

$$-v(\epsilon) := -\left(\frac{1}{3} - \frac{4}{9}\epsilon\right) \quad (41)$$

at the origin  $\mathbf{x} = \mathbf{0}$ .

For the sake of clarity, we may omit obvious variable dependencies in the following analysis.

The rescaled distribution in the unital fragment simply is

$$\widetilde{W}_{\rho_S|\frac{1}{3}\mathbb{1}}(\mathbf{x}) = \frac{W_{\rho_S}(\mathbf{x})}{W_{\frac{1}{3}\mathbb{1}}(\mathbf{x})} = dW_{\rho_S}(\mathbf{x}), \quad (42)$$

<sup>3</sup> Hereinafter, we may drop parameters  $\epsilon$  and  $n$  for clarity

so ordering the rescaled distribution is equivalent to ordering  $W_{\rho_S}$ .

The component values and multiplicities ( $m_i$ ) in the  $n$ -copy case are

$$m_i = 8^{2i} \binom{n}{2i}, \quad (43)$$

$$w(\rho_S)_i = u^{2i} (-v)^{n-2i}, \quad (44)$$

$$w(\mathbb{1}/3)_i = \frac{1}{9^n}, \quad (45)$$

where index  $i$  runs through  $0, \dots, \frac{n}{2}$ . This is derived in Appendix E 2.

It is readily seen that the maximum Wigner component is achieved when  $i = 0$ , so the corresponding multiplicity and Wigner component are  $m_0 = 1$ ,  $w(\rho_S)_0 = v^n$ . The first elbow coordinates therefore can be expressed as

$$(x_0, L_0) = (m_0 w(\mathbb{1}/3)_0, m_0 w(\rho_S)_0) = \left( \frac{1}{9^n}, v^n \right) \quad (46)$$

The first elbow of the target state is located at  $x'_0 = \frac{1}{9^{n'}} > x_0$ , so we can use the first elbow condition,

$$\frac{L_0}{x_0} \geq \frac{L'_0}{x'_0}, \quad (47)$$

to compute analytical distillation bounds for distillation rate  $R = R(\epsilon, \epsilon') := n'/n$  in the unital fragment. We substitute coordinates from Eq. (79) appropriately in Eq. (80) to get

$$R \leq \frac{\ln(3 - 4\epsilon)}{\ln(3 - 4\epsilon')}. \quad (48)$$

Specifically for the problem of distilling pure magic states ( $\epsilon' = 0$ ), we obtain an upper bound in the unital fragment given by

$$R \leq 1 + \frac{\ln(1 - \frac{4}{3}\epsilon)}{\ln 3}. \quad (49)$$

Similar upper bounds on distillation rates for qudits of odd prime dimension include the mana bound [10] and the max-thauma bound [35] which is defined via a semi-definite program, but possesses properties that allow for an easy comparison with our bound. The mana bound can be directly calculated as

$$R \leq \frac{\text{mana}(\rho_S)(\epsilon)}{\text{mana}(\rho_S)(0)} = 1 + \frac{\ln(1 - \frac{8}{15}\epsilon)}{\ln \frac{5}{3}}. \quad (50)$$

Using the properties of super-additivity and vanishing at free states of the max-thauma, we obtain

$$\theta_{\max}(\rho_S(\epsilon)) \geq (1 - \epsilon) |S\rangle\langle S| + \epsilon \frac{1}{3} \mathbb{1} = (1 - \epsilon) \theta_{\max}(\rho_S(0)), \quad (51)$$

so that for a deterministic single-copy distillation process, an initial number of copies  $n \geq \theta_{\max}(\tau)/\theta_{\max}(\rho) \geq 1 - \epsilon$

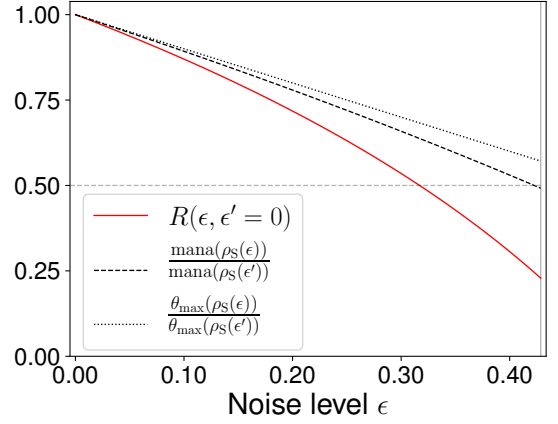


FIG. 5. **Distillation bounds in the unital fragment.** Distillation bounds obtained by majorisation, mana and the line  $1 - \epsilon$  (tightest bound of max-thauma) are plotted for  $\epsilon' = 0$ , up to noise level  $\epsilon = \frac{3}{7}$ . Majorisation consistently provides stricter rates than mana and max-thauma.

is required. Therefore, the max-thauma bound for the pure Strange state distillation process can get as tight as

$$R = \frac{1}{n} = \frac{\theta_{\max}(\rho_S(\epsilon))}{\theta_{\max}(\rho_S(0))} \leq 1 - \epsilon. \quad (52)$$

Both the mana and the max-thauma bounds are looser than the majorization bound we derived via the first elbow constraint as illustrated in Fig. (5).

## V. MAGIC BOUNDS IN ARBITRARY STABILISER FRAGMENTS

We now generalise the approach taken in the previous section and consider bounds on magic distillation for an arbitrary stabiliser fragment,  $\mathcal{R}_\sigma$  where  $\sigma$  is any quantum state  $\sigma \in \text{STAB}$ . In other words, we consider those bounds on distillation that apply when the free operations have  $\sigma$  as a fixed point.

These bounds can be interpreted in two different ways: on one hand they can be viewed as a family of upper bounds parameterized by a stabiliser state  $\sigma$ , on the other we can associate  $\sigma$  to actual hardware limitations or to biased noise models in which it is an equilibrium state of some kind. Without loss of generality, we can always write a stabiliser state  $\sigma$  as a Gibbs state  $\sigma = \gamma_\beta := \frac{1}{Z_\beta} e^{-\beta H}$  for some  $\beta \geq 0$  and Hamiltonian  $H$  as discussed in Section III C.

We focus on the distillation process

$$\rho_S(\epsilon)^{\otimes n} \longrightarrow \rho_S(\epsilon')^{\otimes n'} \otimes \sigma^{\otimes (n-n')}, \quad (53)$$

where the noisy Strange state is given in Eq. (36) and we highlight again that any magic state can be transformed to this form via Clifford operations. Notice that tensoring in copies of  $\gamma_\beta$  does not affect the process, since the state

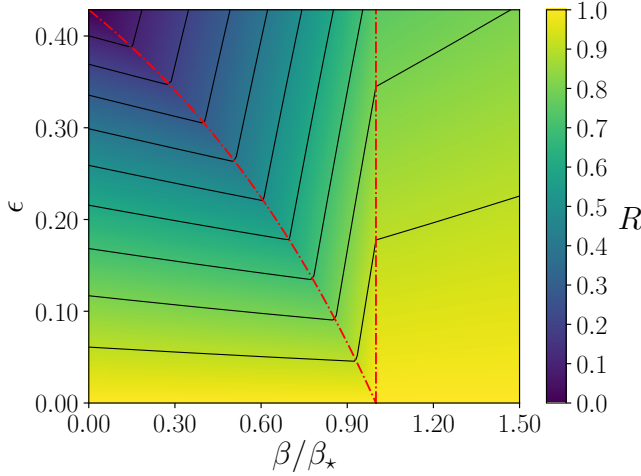


FIG. 6. **Bounds on magic distillation rates  $R(\epsilon, \beta)$  within stabiliser fragments.** The vertical dashed line is the ‘Landauer-like’ temperature threshold  $\beta_*$  and the diagonal dashed curve corresponds to the noise threshold  $\epsilon_*(\beta)$  at every  $\beta \leq \beta_*$ . The unital fragment corresponds to the  $\beta = 0$  line.

is free. In the following result, all copies  $n, n', n - n'$  are even with  $n > n'$ , and  $\epsilon' = 0$ , but the bounds are easily generalised to odd numbers of copies and  $\epsilon'$  such that  $0 < \epsilon' < \epsilon$ . Finally, we assume a range of initial noise levels,  $\epsilon \leq 3/7$ , so that the largest Wigner component of the 1-copy noisy magic state is negative. The numerical value of  $3/7$  is higher than corresponding values of relevant existing distillation protocol error thresholds [28, 31].

Given this context, we now provide the following result on bounding the distillation rate  $R = R(\epsilon, \epsilon', \beta) := \frac{n'}{n}$  of the process in Eq. (53). We also write  $R(\epsilon, \beta) = R(\epsilon, \epsilon' = 0, \beta)$ . The bounds depend on the free energy  $F_\beta$  of state  $\sigma$ ,

$$F_\beta := \text{tr}[H\sigma] - \beta^{-1}S(\sigma) = -\beta^{-1} \log \mathcal{Z}_\beta, \quad (54)$$

where the von Neumann entropy is  $S(\sigma) := -\text{tr}[\sigma \log \sigma]$ .

**Theorem 10.** *Let  $\sigma$  be a qutrit stabiliser state, given by  $\sigma = \frac{1}{\mathcal{Z}} e^{-\beta H}$  with  $H$  having eigenvalues  $E_0 \leq E_1 \leq E_2$ , and where  $\beta$  is an (effective) inverse temperature for the state. We define*

$$\beta_* := \frac{1}{E_2 - E_0} \ln 2 \quad (55)$$

and for  $\beta \leq \beta_*$ , we define a threshold noise level

$$\epsilon_*(\beta) := 3 - \frac{18}{8 - e^{(E_2 - E_0)\beta}}. \quad (56)$$

Then any distillation rate  $R(\epsilon, \beta)$  in the  $\sigma$ -fragment of a magic theory is bounded as follows:

If  $\beta \leq \beta_*$  and  $\epsilon \leq \epsilon_*$ ,

$$R(\epsilon, \beta) \leq 1 + \frac{\ln(1 - \frac{4}{3}\epsilon)}{\beta(E_0 - F_\beta)}. \quad (57)$$

If  $\beta \leq \beta_*$  and  $\epsilon_* < \epsilon$ , then

$$R(\epsilon, \beta) \leq 1 + \frac{\ln(1 - \frac{1}{3}\epsilon) - (E_2 - E_0)(\beta_* - \beta)}{\beta(E_0 - F_\beta)}. \quad (58)$$

Otherwise if  $\beta > \beta_*$ ,

$$R(\epsilon, \beta) \leq 1 + \frac{\ln(1 - \frac{1}{3}\epsilon)}{-\ln 2 + \beta(E_2 - F_\beta)}. \quad (59)$$

A few comments can be given on this result. Firstly, the specific numerical factors in  $\epsilon_*$  are a result of our choice of magic state. Secondly, these bounds are derived based on analysis of only a part of the Lorenz curves and can be improved via a finer analysis. This is apparent by simple numerical calculations on the entirety of the curves. Specifically, the existing bounds follow by considering the dominant terms in the rescaled Wigner distribution and do not, for example, take into account the Lorenz curve’s peak structure. Finally, it is striking that we obtain a Landauer-like condition with a characteristic temperature  $kT_* \ln 2 = E_2 - E_0$ , where  $kT_* = \beta_*^{-1}$ . It is unclear whether this points to a generic feature that can be directly related to fundamental thermodynamic relations, such as the erasure cost of a single bit being  $kT \ln 2$ . Given that we work with qutrits this seems surprising, but does deserve further study.

*Proof.* We provide the full proof here, with some technical details pushed to Appendix F.

Let  $\sigma = e^{-\beta H} / \mathcal{Z}_\beta$  be a stabiliser state, where  $\beta \geq 0$  and  $H = E_0 |0\rangle\langle 0| + E_1 |1\rangle\langle 1| + E_2 |2\rangle\langle 2|$ , with  $E_k \geq 0$ . Its eigen-decomposition can be written as

$$\sigma = \frac{e^{-\beta E_0}}{\mathcal{Z}_\beta} |\varphi_0\rangle\langle \varphi_0| + \frac{e^{-\beta E_1}}{\mathcal{Z}_\beta} |\varphi_1\rangle\langle \varphi_1| + \frac{e^{-\beta E_2}}{\mathcal{Z}_\beta} |\varphi_2\rangle\langle \varphi_2|, \quad (60)$$

where  $\{|\varphi_k\rangle\}$  are pure, orthonormal stabiliser states. Therefore, there exists a Clifford operation that maps  $\sigma$  to  $\gamma_\beta$ , where

$$\gamma_\beta = \frac{e^{-\beta E_0}}{\mathcal{Z}_\beta} |0\rangle\langle 0| + \frac{e^{-\beta E_1}}{\mathcal{Z}_\beta} |1\rangle\langle 1| + \frac{e^{-\beta E_2}}{\mathcal{Z}_\beta} |2\rangle\langle 2|. \quad (61)$$

[Prove this carefully]

This Clifford operation permutes the Hamiltonian eigenvalues on the phase space, so that the negative component  $-v$  of  $\rho_S(\epsilon)$  can lie on the same point on the phase space as any of the eigenvalues. For this reason, we impose no order between the eigenvalues, but simply choose  $E_0$  as the eigenvalue that is associated with the state negativity and denote the highest energy by  $E_{\max} := \max\{E_0, E_1, E_2\}$ .

The Wigner distribution of state  $\gamma_\beta$  can be seen as the ensemble average of the distributions of the compu-

tational basis states,

$$W_{\gamma\beta}(\mathbf{x}) = \sum_{k=0}^2 \frac{e^{-\beta E_k}}{\mathcal{Z}_\beta} W_{|k\rangle\langle k|}(\mathbf{x}) \quad (62)$$

$$= \sum_{k=0}^2 \frac{e^{-\beta E_k}}{\mathcal{Z}_\beta} \delta_{x_0,k} = \frac{e^{-\beta E_{x_0}}}{3\mathcal{Z}_\beta}, \quad (63)$$

for all  $\mathbf{x} \in \mathcal{P}_3$ . All Wigner components are strictly positive, therefore the pre-order  $\prec_{\gamma\beta}$  is always well-defined.

Our aim is to obtain a distillation bound for Eq. (53) which depends on variables  $n, n', \epsilon, \epsilon'$  as well as  $\beta$ . In the analysis that follows, we again drop obvious variable dependencies for clarity.

We construct the Strange state rescaled distribution

$$\widetilde{W}_{\rho_S|\gamma}(\mathbf{x}) = \frac{W_{\rho_S}(\mathbf{x})}{W_\gamma(\mathbf{x})}, \quad (64)$$

which attains four distinct values on the phase space, naturally splitting it into four regions as illustrated in the Fig. (7).

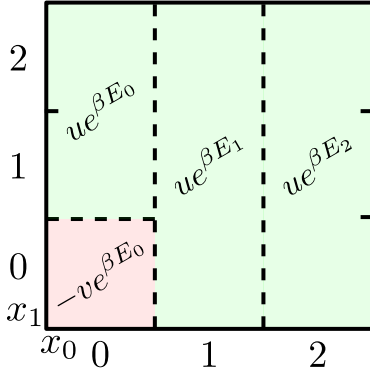


FIG. 7. **Qutrit phase space regions with different rescaled values.** The rescaled distribution attains a unique value in each of the four regions, given by  $3\mathcal{Z} \times$  the value depicted in the region, according to Eq. (68).

The component values and multiplicities of the relevant distributions in the four distinct regions are summarised by the following component and multiplicity vectors,

$$\mathbf{m} := (1, 2, 3, 3), \quad (65)$$

$$\mathbf{w}(\rho_S) := (-v, u, u, u), \quad (66)$$

$$\mathbf{w}(\gamma) := \frac{1}{3\mathcal{Z}} (e^{-\beta E_0}, e^{-\beta E_0}, e^{-\beta E_1}, e^{-\beta E_2}), \quad (67)$$

$$\mathbf{w}(\rho_S|\gamma) := 3\mathcal{Z} (-ve^{\beta E_0}, ue^{\beta E_0}, ue^{\beta E_1}, ue^{\beta E_2}). \quad (68)$$

Using this notation, the component values and multiplicities of the  $n$ -copy distributions can be readily provided by Lemma 22 in Appendix F 1. They are parametrised by three independent components  $i, j, k$ , with sum  $\alpha := i + j + k \leq n$ . Then, the  $n$ -copy multiplicity is given by

$$m_{ijk} = \frac{n!}{i!j!k!(n-\alpha)!} 2^i 3^j 3^k, \quad (69)$$

while the distribution values that correspond to the same index triplet  $(i, j, k)$  are given by

$$w(\rho_S)_{ijk} = (-v)^{n-\alpha} u^\alpha, \quad (70)$$

$$w(\gamma)_{ijk} = (3\mathcal{Z})^{-n} e^{-\beta(n-\alpha)E_0} e^{-\beta(iE_0+jE_1+kE_2)}, \quad (71)$$

$$w(\rho_S|\gamma)_{ijk} = (3\mathcal{Z})^n (-v)^{n-\alpha} u^\alpha e^{\beta(n-\alpha)E_0} e^{\beta(iE_0+jE_1+kE_2)}. \quad (72)$$

In order to construct the  $n$ -copy Lorenz curve  $L_{\rho_S^{\otimes n}|\gamma}$  we need to sort the components of the  $n$ -copy rescaled distribution,  $w(\rho_S|\gamma)_{ijk}$  in decreasing order. In particular, to find the coordinates of the first elbow  $(x_0, L_0)$ , we need to evaluate the maximum rescaled component,

$$\mathbf{w}(\rho_S|\gamma)_{\max} := \quad (73)$$

$$(3\mathcal{Z})^n \max_{i,j,k} \left\{ (-v)^{n-\alpha} u^\alpha e^{\beta(n-\alpha)E_0} e^{\beta(iE_0+jE_1+kE_2)} \right\},$$

where  $0 \leq i, j, k \leq n$  and  $\alpha := i + j + k \leq n$ . Notice that for  $0 \leq \epsilon \leq 3/7$ , we have  $v \geq u$ . We assume that  $n$  is even, so that we need the sum  $\alpha = i + j + k$  to be even for the expression to be positive. The following analysis is similar if  $n$  is chosen to be odd.

Given an even value for the sum  $\alpha$ , the term  $v^{n-\alpha} u^\alpha e^{-\beta(n-\alpha)E_0}$  is fixed, so the expression is maximised by setting the coefficient of the highest energy  $E_{\max}$  equal to  $\alpha$ . Hence, we have

$$\mathbf{w}(\rho_S|\gamma)_{\max} = (3\mathcal{Z})^n v^n e^{n\beta E_0} \max_{\substack{\alpha=0,2, \\ \dots, n-2, n}} \left\{ \left( \frac{u}{v} e^{\beta(E_{\max}-E_0)} \right)^\alpha \right\}. \quad (74)$$

If the expression  $\frac{u}{v} e^{\beta(E_{\max}-E_0)}$  is less than 1 then the maximum occurs at  $\alpha = 0$ , otherwise the maximum occurs at  $\alpha = n$ . To determine this transition we set

$$\frac{u(\epsilon)}{v(\epsilon)} e^{\beta(E_{\max}-E_0)} := \frac{3-\epsilon}{6-8\epsilon} e^{\beta(E_{\max}-E_0)} = 1. \quad (75)$$

We want to find in which cases there exists a threshold noise level  $\epsilon_*$  at which the transition in Eq. (75) occurs. If  $E_{\max} = E_0$ , namely if the state negativity lies in the same phase space region as the highest energy, this threshold is constant in temperature and given by  $\epsilon_* = 3/7$ . Otherwise, there is a threshold temperature value  $\beta_*$  given by

$$\beta_* := \frac{1}{E_{\max} - E_0} \ln 2. \quad (76)$$

Below the threshold,  $0 \leq \beta \leq \beta_*$ , the transition is well defined and the threshold noise level at which it occurs is given by

$$\epsilon_*(\beta) := 3 - \frac{18}{8 - e^{(E_{\max}-E_0)\beta}}. \quad (77)$$

This encompasses the case  $E_{\max} = E_0$ . For  $\beta > \beta_*$ , we have

$$\frac{3-\epsilon}{6-8\epsilon} e^{\beta(E_{\max}-E_0)} > \frac{3-\epsilon}{6-8\epsilon} 2 \geq \frac{1}{2} 2 = 1,$$

so there is no transition and we set  $\epsilon_\star = 0$ .

The maximum rescaled component can then be expressed as

$$\mathbf{w}(\rho_S|\gamma)_{\max} = \begin{cases} (3\mathcal{Z})^n v^n e^{n\beta E_0}, & \epsilon \leq \epsilon_\star, \\ (3\mathcal{Z})^n u^n e^{n\beta E_{\max}}, & \epsilon > \epsilon_\star. \end{cases} \quad (\text{C1})$$

$$(\text{C2})$$

Case (C1) corresponds to  $(i, j, k) = (0, 0, 0)$ , so the multiplicity is  $m_{000} = 1$  and the corresponding Wigner components are  $\mathbf{w}(\rho_S)_{000}, \mathbf{w}(\gamma)_{000}$ . Case (C2) corresponds to

$$(i, j, k) = \begin{cases} (0, n, 0), & \text{if } E_{\max} = E_1, \\ (0, 0, n), & \text{if } E_{\max} = E_2. \end{cases} \quad (78)$$

and we have  $E_{\max} = E_1$  ( $E_{\max} = E_2$ ), so the multiplicity is  $m_{0n0} = 3^n$  ( $m_{00n} = 3^n$ ) and the corresponding Wigner components are  $\mathbf{w}(\rho_S)_{0n0}, \mathbf{w}(\gamma)_{0n0}$  ( $\mathbf{w}(\rho_S)_{00n}, \mathbf{w}(\gamma)_{00n}$ ).

The first elbow coordinates can finally be derived as

$$(x_0, L_0) = \begin{cases} \left( \left( \frac{e^{-\beta E_0}}{3\mathcal{Z}_\beta} \right)^n, v^n \right), & (\text{C1}) \\ \left( \left( \frac{e^{-\beta E_{\max}}}{\mathcal{Z}_\beta} \right)^n, (3u)^n \right). & (\text{C2}) \end{cases} \quad (79)$$

The Lorenz curves of the initial and target states may each be described by either (C1) or (C2), depending on the physical parameters  $\epsilon, \epsilon', \beta$ . Specifically, we have three scenarios:

1. (C1)  $\rightarrow$  (C1) if  $E_{\max} = E_0$  or  $E_{\max} > E_0, \beta < \beta_\star$  and  $\epsilon' < \epsilon \leq \epsilon_\star$ .
2. (C2)  $\rightarrow$  (C1) if  $E_{\max} > E_0, \beta < \beta_\star$  and  $\epsilon' \leq \epsilon_\star < \epsilon$ .
3. (C2)  $\rightarrow$  (C2) if  $E_{\max} > E_0, \beta < \beta_\star$  and  $\epsilon_\star \leq \epsilon' < \epsilon$  or  $E_{\max} > E_0, \beta \geq \beta_\star$ .

Note that (C1)  $\rightarrow$  (C2) is impossible because it would imply  $\beta < \beta_\star$  and  $\epsilon \leq \epsilon_\star \leq \epsilon' < \epsilon$ , a contradiction.

In all three scenarios, it is simple to check that the initial state's first elbow is always located to the left (closer to 0) of the target state's first elbow,  $x_0 \leq x'_0$ , as proven

in Appendix F 2, where the prime ( $'$ ) is used to indicate target state coordinates. Therefore, we can use the first elbow condition,

$$\frac{L_0}{x_0} \geq \frac{L'_0}{x'_0}, \quad (80)$$

to compute analytical distillation bounds for the distillation rate  $R = R(\epsilon, \epsilon', \beta) := n'/n$  in all three possible scenarios. Involving more elbows gives stricter, but more convoluted necessary distillation constraints.

We substitute coordinates from Eq. (79) appropriately in Eq. (80) to get the following necessary conditions,

$$R \leq \begin{cases} \frac{\ln(1 - \frac{4}{3}\epsilon) + \beta(E_0 - F_\beta)}{\ln(1 - \frac{4}{3}\epsilon') + \beta(E_0 - F_\beta)}, & (\text{C1}) \rightarrow (\text{C1}), \\ \frac{\ln(\frac{1}{2} - \frac{1}{6}\epsilon) + \beta(E_{\max} - F_\beta)}{\ln(1 - \frac{4}{3}\epsilon') + \beta(E_0 - F_\beta)}, & (\text{C2}) \rightarrow (\text{C1}), \\ \frac{\ln(\frac{1}{2} - \frac{1}{6}\epsilon) + \beta(E_{\max} - F_\beta)}{\ln(\frac{1}{2} - \frac{1}{6}\epsilon') + \beta(E_{\max} - F_\beta)}, & (\text{C2}) \rightarrow (\text{C2}). \end{cases} \quad (81)$$

Equivalently, we require that

$$\epsilon \leq \begin{cases} \frac{3}{4} - \frac{3}{4} \left(1 - \frac{4}{3}\epsilon'\right)^R \left(\frac{e^{-\beta E_0}}{\mathcal{Z}_\beta}\right)^{1-R}, & (\text{C1}) \rightarrow (\text{C1}), \\ 3 - 6 \left(1 - \frac{4}{3}\epsilon'\right)^R \frac{e^{-\beta E_{\max}} e^{R\beta E_0}}{\mathcal{Z}_\beta^{1-R}}, & (\text{C2}) \rightarrow (\text{C1}), \\ 3 - 6 \left(\frac{1}{2} - \frac{1}{6}\epsilon'\right)^R \left(\frac{e^{-\beta E_{\max}}}{\mathcal{Z}_\beta}\right)^{1-R}, & (\text{C2}) \rightarrow (\text{C2}). \end{cases} \quad (82)$$

Substituting  $\epsilon' = 0$  in Eq. (81) leads to the theorem statement.  $\square$

## VI. CONCLUSION

[Summary]

- 
- [1] R. Raussendorf and H. J. Briegel, *Phys. Rev. Lett.* **86**, 5188 (2001).  
[2] D. Gross and J. Eisert, *Phys. Rev. Lett.* **98**, 220503 (2007).  
[3] I. L. Markov and Y. Shi, *SIAM J. Comput.* **38**, 963 (2008).  
[4] D. Gross, S. T. Flammia, and J. Eisert, *Phys. Rev. Lett.* **102**, 190501 (2009).  
[5] M. Van den Nest, A. Miyake, W. Dür, and H. J. Briegel, *Phys. Rev. Lett.* **97**, 150504 (2006).  
[6] M. Van den Nest, W. Dür, G. Vidal, and H. J. Briegel, *Phys. Rev. A* **75**, 012337 (2007).  
[7] G. Vidal, *Phys. Rev. Lett.* **98**, 070201 (2007).  
[8] K. Fujii and M. Hayashi, *Phys. Rev. A* **96**, 030301 (2017).  
[9] D. Gottesman, *Stabilizer codes and quantum error correction*, Ph.D. thesis, California Institute of Technology (1997).  
[10] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, *New Journal of Physics* **14**, 113011 (2012).  
[11] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson, *New Journal of Physics* **16**, 013009 (2014).



- [12] X. Wang, M. M. Wilde, and Y. Su, arXiv e-prints , arXiv:1903.04483 (2019), [arXiv:1903.04483](#).
- [13] X. Wang, M. M. Wilde, and Y. Su, arXiv e-prints , arXiv:1812.10145 (2018), [arXiv:1812.10145 \[quant-ph\]](#).
- [14] M. Howard and E. Campbell, *Phys. Rev. Lett.* **118**, 090501 (2017).
- [15] E. T. Campbell and D. E. Browne, *Phys. Rev. Lett.* **104**, 030503 (2010).
- [16] D. Gross, *Journal of Mathematical Physics* **47**, 122107 (2006).
- [17] D. Gross, *Applied Physics B* **86**, 367 (2007).
- [18] P. Źwikliński, M. Studziński, M. Horodecki, and J. Oppenheim, *Phys. Rev. Lett.* **115**, 210403 (2015).
- [19] M. Lostaglio, D. Jennings, and T. Rudolph, *Nature Communications* **6** (2015), 10.1038/ncomms7383.
- [20] M. Lostaglio, arXiv e-prints , arXiv:1807.11549 (2018).
- [21] G. Gour, D. Jennings, F. Buscemi, R. Duan, and I. Marvian, *Nature Communications* **5352** (2018), 10.1038/s41467-018-06261-7.
- [22] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and T. Beth, *International Journal of Theoretical Physics* **39**, 2717 (2000).
- [23] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, *Phys. Rev. Lett.* **111**, 250404 (2013).
- [24] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Yunger Halpern, *Physics Reports* **583** (2015), 0.1016/j.physrep.2015.04.003.
- [25] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications* (Springer, 2011).
- [26] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, United Kingdom, 2010).
- [27] M. Ahmadi, H. B. Dang, G. Gour, and B. C. Sanders, *Phys. Rev. A* **97**, 062332 (2018).
- [28] S. Bravyi and A. Kitaev, *Phys. Rev. A* **71**, 022316 (2005).
- [29] A. Mari and J. Eisert, *Phys. Rev. Lett.* **109**, 230503 (2012).
- [30] M. Howard, J. Wallman, V. Veitch, and J. Emerson, *Nature* **510** (2014), 10.1038/nature13460.
- [31] S. Prakash, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **476**, 20200187 (2020), <https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.2020.0187>.
- [32] A. Jain and S. Prakash, *Phys. Rev. A* **102**, 042409 (2020).
- [33] J. R. Seddon, B. Regula, H. Pashayan, Y. Ouyang, and E. T. Campbell, “Quantifying quantum speedups: improved classical simulation from tighter magic monotones,” (2020), [arXiv:2002.06181 \[quant-ph\]](#).
- [34] K. Fang and Z.-W. Liu, *Phys. Rev. Lett.* **125**, 060405 (2020).
- [35] X. Wang, M. M. Wilde, and Y. Su, *Phys. Rev. Lett.* **124**, 090505 (2020).
- [36] R. Ash, *Information Theory* (Dover Publications Inc., 1965).

## Appendix A: Properties of Wigner distributions

Here we present important properties of the Wigner distribution that are used throughout the paper.

**Proposition 11.** *The Wigner distribution of a state  $\rho \in \mathcal{B}(\mathcal{H}_d)$  is*

- (i) *Real valued:*  $W_\rho \in \mathbb{R}^{d^2}$ ;
- (ii) *Normalised:*  $\sum_{\mathbf{z} \in \mathcal{P}_d} W_\rho(\mathbf{z}) = 1$ ;
- (iii) *Bounded:*  $|W_\rho(\mathbf{x})| \leq \frac{1}{d}$ .
- (iv) *Additive under mixing:*  

$$W_{\sum_i p_i \rho_i}(\mathbf{x}) = \sum_i p_i W_{\rho_i}(\mathbf{x});$$
- (v) *Multiplicative under tensor products:*  

$$W_{\rho_A \otimes \rho_B}(\mathbf{x}_A \oplus \mathbf{x}_B) = W_{\rho_A}(\mathbf{x}_A) W_{\rho_B}(\mathbf{x}_B).$$

Not all distributions satisfying these properties correspond to positive semi-definite states.

*Proof.* [Proof?] □

**Proposition 12.** *The Wigner distribution of a CPTP operation  $\mathcal{E} : \mathcal{B}(\mathcal{H}_{d_A}) \mapsto \mathcal{B}(\mathcal{H}_{d_B})$  is*

- 1. *Real-valued:*  $W_{\mathcal{E}} \in \mathbb{R}^{d^2} \times \mathbb{R}^{d^2}$ ;
- 2. *Normalised:*  $\sum_{\mathbf{z} \in \mathcal{P}_{d_B}} W_{\mathcal{E}}(\mathbf{z}|\mathbf{x}) = 1$  for any  $\mathbf{x} \in \mathcal{P}_{d_A}$ ;
- 3. *Bounded:*  $|W_{\mathcal{E}}(\mathbf{y}|\mathbf{x})| \leq \frac{d_A}{d_B}$ ;
- 4. *Transitive:*  $W_{\mathcal{E}(\rho)}(\mathbf{y}) = \sum_{\mathbf{z} \in \mathcal{P}_{d_A}} W_{\mathcal{E}}(\mathbf{y}|\mathbf{z}) W_\rho(\mathbf{z})$  for any  $\mathbf{y} \in \mathcal{P}_{d_B}$ .

If  $d_A = d_B$ , and in particular if operation  $\mathcal{E}$  maps a Hilbert space onto itself, then the stochasticity condition  $|W_{\mathcal{E}}(\mathbf{y}|\mathbf{x})| \leq 1$  is satisfied.

*Proof.* [Proof?] □

A  $(d \times d)$  matrix  $S$  is in  $S_d(\mathbf{d})$ , if

- 1.  $S_{ij} \geq 0$  for all  $i, j \in \mathbb{Z}_d$ ;
- 2.  $\sum_{j=1}^n S_{ij} = 1$  for all  $i \in \mathbb{Z}_d$ ;
- 3.  $S\mathbf{d} = \mathbf{d}$ .

## Appendix B: Properties of majorization

### 1. Simple-majorization equivalence conditions

In the unital fragment, namely the limit of infinite temperature,  $\beta = 0$ , the free state is the maximally mixed state  $\frac{1}{d} \mathbb{1}$  with uniform Wigner distribution  $\frac{1}{d} \mathbf{1} = (\frac{1}{d}, \dots, \frac{1}{d})$ . This fragment is governed by simple majorization and we first prove strong equivalences for this type of majorization.

**Theorem 13.** *Given  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  the following statements are equivalent:*

1.  $\mathbf{x} \prec \mathbf{y}$ ;
2.  $\mathbf{x} = B\mathbf{y}$  for bistochastic  $B$ ;
3.  $\sum_{i=1}^n |x_i - t| \leq \sum_{i=1}^n |y_i - t|$  for all  $t \in \mathbb{R}$ ;
4.  $\sum_{i=1}^n (x_i - t)^+ \leq \sum_{i=1}^n (y_i - t)^+$  for all  $t \in \mathbb{R}$  and  $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ , where  $(x)^+ = \max\{x, 0\}$ ;
5.  $L_{\mathbf{x}}(k) \leq L_{\mathbf{y}}(k)$  for  $k = 1, \dots, n-1$  and  $L_{\mathbf{x}}(n) = L_{\mathbf{y}}(n)$ .

*Proof.* [Proof?]  $\square$

## 2. Embedding map

Any  $\mathbf{d}$ -majorization problem can be rephrased as a simple majorization problem in a higher dimensional space via the embedding map.

**Definition 14.** The embedding map  $\Gamma_{\mathbf{d}} : \mathbb{R}^n \mapsto \mathbb{R}^D$ ,  $D = \sum_{i=1}^n d_i$  is the function

$$\Gamma_{\mathbf{d}}(\mathbf{w}) = \bigoplus_{i=1}^n w_i \frac{1}{d_i} \mathbf{1}, \quad (\text{B1})$$

where  $\mathbf{1}/d_i$  is the  $d_i$ -dimensional uniform distribution. The left inverse  $\Gamma_{\mathbf{d}}^{-1} : \mathbb{R}^D \mapsto \mathbb{R}^n$  is defined to sum up the elements in each block of  $\Gamma_{\mathbf{d}}(\mathbf{w})$ , so that

$$\Gamma_{\mathbf{d}}^{-1}(\bigoplus_{i=1}^n w_i \mathbf{1}/d_i) = \mathbf{w}. \quad (\text{B2})$$

This is not a right inverse, because  $\Gamma_{\mathbf{d}}$  is not surjective.

The direct sum simply amounts to listing the uniform distributions one after the other. The embedding map maps the Gibbs distribution to the uniform distribution,  $\Gamma_{\mathbf{d}}(\mathbf{d}) = \mathbf{1}/D$ . Then, a non-increasing ordering  $\Gamma_{\mathbf{d}}(\mathbf{z})^\downarrow$  in the new space, corresponds to the so-called “ $\beta$ -ordering” of the original vector denoted by the permutation  $\pi$  in Definition 4, mapping  $(w_i/d_i) \mapsto (w_i/d_i)^\downarrow$  for all  $i = 1, \dots, n$ .

## 3. $\mathbf{d}$ -majorization equivalence conditions

**Theorem 15.** Given  $\mathbf{x}, \mathbf{y}, \mathbf{d} \in \mathbb{R}^n$ , such that the components of  $\mathbf{d}$  are positive, the following statements are equivalent:

1.  $\mathbf{x} \prec_{\mathbf{d}} \mathbf{y}$ ;
2.  $\Gamma_{\mathbf{d}}(\mathbf{x}) \prec \Gamma_{\mathbf{d}}(\mathbf{y})$ ;
3.  $\sum_{i=1}^n |x_i - td_i| \leq \sum_{i=1}^n |y_i - td_i|$  for all  $t \in \mathbb{R}$ ;

4.  $\sum_{i=1}^n (x_i - td_i)^+ \leq \sum_{i=1}^n (y_i - td_i)^+$  for all  $t \in \mathbb{R}$  and  $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ ;
5.  $L_{\mathbf{x}|\mathbf{d}}(k) \leq L_{\mathbf{y}|\mathbf{d}}(k)$  for  $k = 1, \dots, n-1$  and  $L_{\mathbf{x}|\mathbf{d}}(n) = L_{\mathbf{y}|\mathbf{d}}(n)$ .

*Proof.* [Proof?]  $\square$

## Appendix C: Technical properties of $\sigma$ -fragments

[this section hasn't been refereed in various parts of the main text]

**Theorem 16.** Let  $\mathcal{R}_\sigma = (\mathcal{O}_\sigma, \mathcal{F}_\sigma)$  be a  $\sigma$ -fragment of magic theory  $\mathcal{R} = (\mathcal{O}, \mathcal{F})$ . The following statements hold:

1. No  $\sigma$ -fragment is empty.
2. If a free operation leaves two states invariant, then it also leaves their mixtures invariant,

$$\mathcal{O}_\sigma \cap \mathcal{O}_{\sigma'} \subseteq \mathcal{O}_{p\sigma + (1-p)\sigma'} \text{ for all } p \in [0, 1].$$

3. Let  $\mathcal{E}$  be a CPTP operation with Wigner distribution  $W_{\mathcal{E}}$ . For  $\mathcal{R} = \mathcal{R}_{\max}$ ,  $\mathcal{E} \in \mathcal{O}_\sigma$  iff  $W_{\mathcal{E}} \in S_{d^2}(W_\sigma)$ . [I don't like this one]

*Proof.*

1. The identity channel  $1_{\mathcal{D}} : \mathcal{D} \mapsto \mathcal{D}$  belongs to every  $\sigma$ -fragment, as  $1_{\mathcal{D}} \in \mathcal{O}$  and  $1_{\mathcal{D}}\sigma = \sigma$  for all  $\sigma \in \mathcal{F}$ .
2. Let  $\mathcal{E} \in \mathcal{O}_\sigma \cap \mathcal{O}_{\sigma'}$ . Then  $\mathcal{E} \in \text{CPTP}$  and corresponds to stochastic Wigner distribution  $W_{\mathcal{E}}$  such that  $W_{\mathcal{E}}W_\sigma = W_\sigma$  and  $W_{\mathcal{E}}W_{\sigma'} = W_{\sigma'}$ . Then,  $W_{\mathcal{E}}W_{p\sigma + (1-p)\sigma'} = W_{p\sigma + (1-p)\sigma'}$  for any  $p \in [0, 1]$  due to the additive property 11 of the Wigner distribution, implying that state  $p\sigma + (1-p)\sigma'$  is also left invariant by  $\mathcal{E}$ .
3. Let  $\mathcal{O}'_\sigma := \{\mathcal{E} \in \text{CPTP} : W_{\mathcal{E}} \in S_{d^2}(W_\sigma)\}$  be the described set of operations.

Suppose  $\mathcal{E}$  is in  $\mathcal{O}_\sigma$ , then  $\mathcal{E} \in \text{CPTP}$  and  $W_{\mathcal{E}} \in S_{d^2}(W_\sigma)$  due to property 4 of Proposition 12, hence  $\mathcal{O}_\sigma \subseteq \mathcal{O}'_\sigma$ .

Conversely, suppose  $\mathcal{E} \in \text{CPTP}$  with  $W_{\mathcal{E}} \in S_{d^2}(W_\sigma)$ . Then,  $W_{\mathcal{E}}(\mathbf{y}|\mathbf{x}) \geq 0$  for all  $\mathbf{x}, \mathbf{y}$ , hence  $\mathcal{E} \in \mathcal{O}$ . Furthermore,  $W_{\mathcal{E}}W_\sigma = W_\sigma$  implies  $\mathcal{E}(\sigma) = \sigma$  using Eq. (14) defined for any CPTP operation  $\mathcal{E}$ . Hence,  $\mathcal{O}'_\sigma \subseteq \mathcal{O}_\sigma$ .

$\square$

Any free state  $\sigma \in \mathcal{F}$  corresponds to a  $d^2$ -dimensional probability distribution  $W_\sigma$  and any free operation  $\mathcal{E} \in \mathcal{O}$  corresponds to a  $d^2 \times d^2$  stochastic matrix (or conditional probability distribution)  $W_{\mathcal{E}}$ . Note that these mappings

are one-to-one due to the orthogonality of the phase-point operators as an operator basis.

Note further that free states  $\mathcal{F}$  are mapped onto a *strict subset* of the set of probability distributions. As a counterexample, the sharp  $d^2$ -dimensional probability distribution  $(1, 0, \dots, 0)$  does not correspond to any qudit Wigner distribution because of the boundedness condition in Proposition 11.

Similarly, not all stochastic matrices correspond to completely positive operations. As an example, consider the permutation matrix

$$\Pi_X = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \in S_5(W_{\frac{1}{5}\mathbb{1}}). \quad (C1)$$

It preserves the uniform distribution  $W_{\frac{1}{5}\mathbb{1}}$ , but it does not correspond to any positive (hence quantum) operation.

#### Appendix D: Properties of Lorenz curves

[Prove statements underneath Definition 4?]

##### 1. Basic properties

Furthermore, single-copy Lorenz curves are simply additive in noise,

$$L_{(1-\epsilon)\rho + \epsilon\sigma} = (1-\epsilon)L_{\rho} + \epsilon L_{\sigma}, \quad (D1)$$

for any  $\sigma, \rho$ . This is not true for higher number of copies.

##### 2. $\sigma$ -fragment continuity

**Proposition 17.** *The Lorenz curve  $L_{\rho|\sigma}(x)$  of a state  $\rho$  with respect to state  $\sigma$  is uniformly continuous in  $\sigma$ .*

*Proof.* [Idea:] If  $\sigma$  and  $\sigma'$  are  $\delta$ -close with respect to a state norm, then the curves  $L_{\rho|\sigma}(x)$  and  $L_{\rho|\sigma'}(x)$  are  $\epsilon$ -close at all  $x \in [0, 1]$ .  $\square$

##### 3. Area monotone

[I don't believe in the area monotone - should we keep?]

Let  $L_{>1}$  be the set of points on the Lorenz curve  $L_{\rho|\sigma}(k)$  that lie above 1 of state  $\rho$  in the  $\sigma$ -fragment. If  $\rho$  is a free state,  $L_{>1}$  is empty and  $\mathcal{A}_\sigma(\rho) = 0$ . Otherwise,  $\mathcal{A}_\sigma(\rho) > 0$  and it can be calculated exactly using the trapezium rule or the shoelace formula. Let  $k$  be the index of the first point  $(x_k, L_{\rho|\sigma}(k))$  lying above 1. Then  $L_{\rho}(k)$  crosses 1 at

$$x_{\text{int}} = x_k - \frac{x_k - x_{k-1}}{L_{\rho|\sigma}(k) - L_{\rho|\sigma}(k-1)} L_{\rho|\sigma}(k), \quad (D2)$$

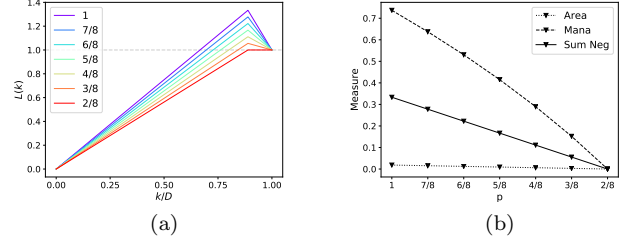


FIG. 8. (a) Lorenz curve of  $p|S\rangle\langle S| + (1-p)\frac{1}{d}\mathbb{1}$  for  $p$  given in the legend. (b) Different measures for the states on the left. [can replace Fig. (3) with a more informative version of this figure]

as well as at  $(x_d, L_{\rho|\sigma}(d)) = (1, 1)$

Now we can define

$$L_{>1}^+ = \{(x_i, y_i)\}_{0 \leq i \leq n} \quad (D3)$$

such that it contains the initial point of intersection  $(x_0, y_0) \equiv (x_{n+1}, y_{n+1}) := (x_{\text{int}}, 1)$ , all points  $(x_i, y_i)$ , labelled by  $i = 1, \dots, n-1$  that lie above  $y = 1$  contained in  $L_{>1}$  and finally the second point of intersection  $(x_n, y_n) = (1, 1)$ . Then,

$$\mathcal{A}_\sigma(\rho) = \frac{1}{2} \sum_{i=0}^n (x_{i+1}y_i - x_iy_{i+1}). \quad (D4)$$

##### 4. Lorenz curve independent constraints

**Theorem 18.** *Let  $\rho, \tau$  be two quantum states with Lorenz curves  $L_{\rho|\sigma}(x), L_{\tau|\sigma}(x)$  in the  $\sigma$ -fragment.*

*Let  $t$  be the number of elbows of  $L_{\tau|\sigma}(x)$  at locations  $x_1, \dots, x_t$ .*

*Then,  $L_{\rho|\sigma}(x) \geq L_{\tau|\sigma}(x)$  for all  $x \in [0, 1]$  iff  $L_{\rho|\sigma}(x_i) \geq L_{\tau|\sigma}(x_i)$  for all  $i = 1, \dots, t$ .*

*Proof.*  $L_{\rho|\sigma}(x) \geq L_{\tau|\sigma}(x)$  for all  $x \in [0, 1]$  trivially implies  $L_{\rho|\sigma}(x_i) \geq L_{\tau|\sigma}(x_i)$  for all  $i = 1, \dots, n'$ .

Conversely, assume that  $L_{\rho|\sigma}(x_i) \geq L_{\tau|\sigma}(x_i)$  for all  $i = 1, \dots, r$ . First, let  $x_0 = 0$  and  $x_{n'+1} = 1$ , so that  $L_{\rho|\sigma}(x_0) = L_{\tau|\sigma}(x_0) = 0$  and  $L_{\rho|\sigma}(x_{n'+1}) = L_{\tau|\sigma}(x_{n'+1}) = 1$ . Hence, we can extend the set of elbows  $E$  to  $E' = E \cup \{x_0, x_{n'+1}\}$ .

Pick two consecutive locations  $x_i, x_{i+1}$  in  $E'$  and consider the line segment  $\ell_\tau(x)$  connecting points  $(x_i, L_{\tau|\sigma}(x_i))$  and  $(x_{i+1}, L_{\tau|\sigma}(x_{i+1}))$  as well as the line segment  $\ell_\rho(x)$  connecting points  $(x_i, L_{\rho|\sigma}(x_i))$  and  $(x_{i+1}, L_{\rho|\sigma}(x_{i+1}))$ . This is illustrated in Fig. (9).

Due to concavity of  $L_{\rho|\sigma}$ , it is clear that for all  $x \in [x_i, x_{i+1}]$ , we have  $L_{\rho|\sigma}(x) \geq \ell_\rho(x) \geq \ell_\tau(x) = L_{\tau|\sigma}(x)$ . This argument can be made in all intervals  $[x_i, x_{i+1}]$  with  $i = 0, \dots, n'$ , so the proof is complete.  $\square$

This theorem is of practical importance in calculating the necessary distillation constraints derived via majorization.

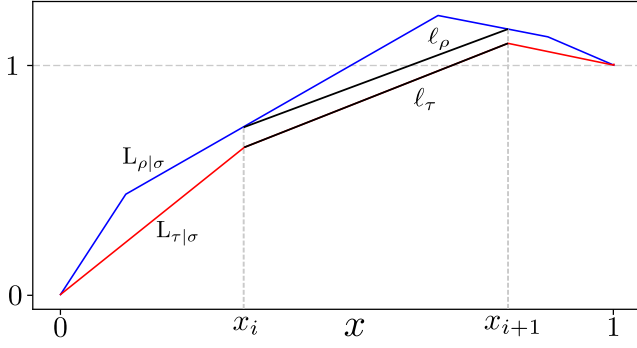


FIG. 9. Illustration of Theorem 18.

[In particular], it can be used in the discussion of standard Lorenz curves to set the necessary constraints,

$$\begin{aligned} L(x_{I'(0,m'_0,9^{n-n'})}) &\geq L(x_{I'(0,m'_0,9^{n-n'})}), \\ L(x_{I'(1,m'_1,9^{n-n'})}) &\geq L(x_{I'(1,m'_1,9^{n-n'})}), \\ &\vdots \\ L(x_{I'(n',m'_{n'},9^{n-n'})}) &\geq L(x_{I'(n',m'_{n'},9^{n-n'})}), \end{aligned}$$

where  $x_{I'(0,m'_0,9^{n-n'})}, x_{I'(1,m'_1,9^{n-n'})}, \dots, x_{I'(n',m'_{n'},9^{n-n'})} \in T$  are the locations of the elbows of  $L_{\rho_S(\epsilon')^{\otimes n'}|\frac{1}{d}\mathbb{1}}$  (excluding 0 and 1).

## Appendix E: Technical derivation for Lorenz curves in unital fragment

### 1. Binomial distributions and error bounds

Consider an experiment consisting of  $n$  trials of throwing a  $p$ -coin, that is a coin with probability  $p$  of landing on one side and  $1-p$  of landing on the other. This scenario describes the construction of an  $n$ -copy Strange state Wigner distribution from a single-copy as presented in ???. The experiment is described by binomial statistics and we write down the sum over an even number of successful trials  $\Phi_+$  (over an odd number of successful trials  $\Phi_-$ ),

$$\Phi_+(af; nf, p) := \sum_{\ell=0}^{af/2} \binom{nf}{2\ell} p^{2\ell} (1-p)^{nf-2\ell},$$

for even integers  $af \in [0, nf]$ , (E1)

$$\Phi_-(af; nf, p) := \sum_{\ell=1}^{(af-1)/2} \binom{nf}{2\ell+1} p^{2\ell+1} (1-p)^{nf-(2\ell+1)},$$

for odd integers  $af \in [0, nf]$ . (E2)

Note that index  $af$  only takes even (odd) values when labelling  $\Phi_+$  ( $\Phi_-$ ). In Appendix E 2, we will use  $\Phi_+$  and

$\Phi_-$  to express the elbow coordinates of standard Lorenz curves.

We define the classical entropy of a  $p$ -coin as well as the classical relative entropy between a  $p$ -coin and a  $q$ -coin,

$$S(p) := -p \log p - (1-p) \log(1-p), \quad (\text{E3})$$

$$S(p||q) := p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}. \quad (\text{E4})$$

They are symmetric in the sense that  $S(p) = S(1-p)$  and  $S(p||q) = S(1-p||1-q)$ .

A useful result is the entropic bound on a combination [36].

**Lemma 19.** For all  $\ell \in [1, np-1]$ ,

$$\left[ 8\ell \left( 1 - \frac{\ell}{np} \right) \right]^{-\frac{1}{2}} 2^{nS(\frac{\ell}{np})} \leq \binom{np}{\ell} \leq \quad (\text{E5})$$

$$\left[ 2\pi\ell \left( 1 - \frac{\ell}{np} \right) \right]^{-\frac{1}{2}} 2^{nS(\frac{\ell}{np})}. \quad (\text{E6})$$

*Proof.* For  $\ell = 1, 2, np-1, np-2$  check by direct calculation. For all other cases, use Stirling's approximation. [CITE]  $\square$

With the help of this lemma, we directly arrive at

**Theorem 20.** Given fixed  $n > 0$  and  $p$ ,  $\Phi_+, \Phi_-$  satisfy the following bounds:

1.  $\Phi_+(a; n, p) \geq \sum_{\ell=0}^{np/2} \left[ 16\ell \left( 1 - \frac{2\ell}{np} \right) \right]^{-\frac{1}{2}} 2^{-nS(\frac{2\ell}{np}||p)},$   
for all even  $a \in [2, np]$
2.  $\Phi_+(a; n, p) \leq \sum_{\ell=0}^{np/2} \left[ 4\pi\ell \left( 1 - \frac{2\ell}{nf} \right) \right]^{-\frac{1}{2}} 2^{-nS(\frac{2\ell}{np}||p)},$   
for all even  $a \in [2, np]$
3.  $\Phi_-(a; n, p) \geq \sum_{\ell=1}^{(np-1)/2} \left[ 16(\ell+1) \left( 1 - \frac{2\ell+1}{np} \right) \right]^{-\frac{1}{2}} \times$   
 $\times 2^{-nS(\frac{2\ell+1}{np}||p)},$  for all odd  $a \in [1, np]$
4.  $\Phi_-(a; n, p) \leq \sum_{\ell=1}^{(np-1)/2} \left[ 4\pi(\ell+1) \left( 1 - \frac{2\ell+1}{nf} \right) \right]^{-\frac{1}{2}} \times$   
 $\times 2^{-nS(\frac{2\ell+1}{np}||p)},$  for all odd  $a \in [1, np]$

*Proof.* All four statements follow from application of Lemma 19 on the combinatorial coefficient and the definition of relative entropy given in Eq. (E4)  $\square$

### 2. Standard Lorenz curve elbow coordinates

The Wigner distribution of the  $n$ -copy qutrit maximally mixed state  $(\mathbb{1}/3)^{\otimes n}$  is the uniform distribution

$$W_{(\mathbb{1}/3)^{\otimes n}} = \left( \overbrace{\frac{1}{9^n}, \dots, \frac{1}{9^n}}^{9^n} \right). \quad (\text{E7})$$

The Wigner distribution of the 1-copy  $\epsilon$ -noisy Strange state  $\rho_S(\epsilon)$  in the unital fragment is a permutation of

$$W_{\rho_S(\epsilon)} = \left( \overbrace{\frac{1}{6} - \frac{1}{18}\epsilon, \dots, \frac{1}{6} - \frac{1}{18}\epsilon}^8, \overbrace{-\frac{1}{3} + \frac{4}{9}\epsilon}^1 \right) \quad (\text{E8})$$

The two distinct components are plotted Fig. (10) as a function of noise. It is clear that in the unital fragment the Strange state contains Wigner negativities in the regime  $0 \leq \epsilon \leq \frac{3}{4}$ .

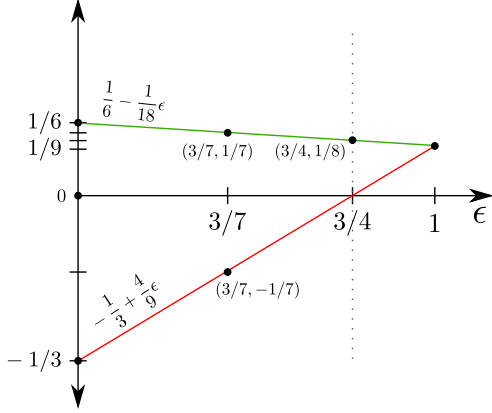


FIG. 10. **Wigner components of the noisy Strange state.** In the interval  $0 \leq \epsilon < 3/7$ , the negative component is larger than the positive components. At  $\epsilon = 3/7$  the Wigner distribution is  $(-\frac{1}{7}, \frac{1}{7}, \dots, \frac{1}{7})$ . In the interval  $3/7 < \epsilon < 3/4$ , the positive components are larger than the negative component. In the interval  $3/4 \leq \epsilon \leq 1$ , there is no negativity.

The Wigner distribution of the  $n$ -copy  $\epsilon$ -noisy Strange state  $\rho_S(\epsilon)^{\otimes n}$  in the unital fragment is given by the convolution  $W_{\rho_S(\epsilon)^{\otimes n}} = W_{\rho_S(\epsilon)}^{\otimes n}$ . In general,  $\rho_S(\epsilon)^{\otimes n}$  contains  $n+1$  distinct components, labelled  $0, \dots, n$ . We present the distinct Wigner components of  $\rho_S(\epsilon)^{\otimes n}$  along with their multiplicities in Table I. Note that LHS (RHS) refers to elbow coordinates  $i$  on the left of and including (right of) the maximum, precisely

$$\text{LHS: } 0 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor; \quad (\text{E9})$$

$$\text{RHS: } \left\lfloor \frac{n}{2} \right\rfloor + 1 \leq i \leq n. \quad (\text{E10})$$

For example, the distribution of state  $\rho_S(0)^{\otimes 2}$  is

$$W_{\rho_S(0)^{\otimes 2}} = \left( \overbrace{\left(-\frac{1}{3}\right)^2}^1, \overbrace{\left(\frac{1}{6}\right)^2, \dots, \left(\frac{1}{6}\right)^2}^{64}, \overbrace{-\frac{1}{3} \cdot \frac{1}{6}, \dots, -\frac{1}{3} \cdot \frac{1}{6}}^{16} \right)$$

Every standard Lorenz curve contains  $n$  elbows, labelled by

$$\{(x_i, L_i)\}_{i=-1,0,\dots,n},$$

Case		$m_i(n, \epsilon)$	$w_i(n, \epsilon)$
$0 \leq \epsilon < \frac{3}{4}$	$n$ even	LHS: $8^{2i} \binom{n}{2i}$ RHS: $8^{n-2i} \binom{n}{2i}$	$\left(\frac{1}{6} - \frac{1}{18}\epsilon\right)^{2i} \left(-\frac{1}{3} + \frac{4}{9}\epsilon\right)^{n-2i}$ $\left(\frac{1}{6} - \frac{1}{18}\epsilon\right)^{n-2i} \left(-\frac{1}{3} + \frac{4}{9}\epsilon\right)^{2i}$
	$n$ odd	LHS: $8^{2i+1} \binom{n}{2i+1}$ RHS: $8^{n-2i-1} \binom{n}{2i+1}$	$\left(\frac{1}{6} - \frac{1}{18}\epsilon\right)^{2i+1} \left(-\frac{1}{3} + \frac{4}{9}\epsilon\right)^{n-2i-1}$ $\left(\frac{1}{6} - \frac{1}{18}\epsilon\right)^{n-2i-1} \left(-\frac{1}{3} + \frac{4}{9}\epsilon\right)^{2i+1}$
	$\frac{3}{4} \leq \epsilon < \frac{3}{2}$	$n$ even: LHS: $8^{n-2i} \binom{n}{2i}$ , RHS: $8^{2i} \binom{n}{2i}$	$\left(\frac{1}{6} - \frac{1}{18}\epsilon\right)^{n-2i} \left(-\frac{1}{3} + \frac{4}{9}\epsilon\right)^{2i}$ $\left(\frac{1}{6} - \frac{1}{18}\epsilon\right)^{2i} \left(-\frac{1}{3} + \frac{4}{9}\epsilon\right)^{n-2i}$
	$\frac{3}{2} \leq \epsilon$	$n$ odd: LHS: $8^{n-2i} \binom{n}{2i}$ , RHS: $8^{2i} \binom{n}{2i}$	$\left(\frac{1}{6} - \frac{1}{18}\epsilon\right)^{n-2i} \left(-\frac{1}{3} + \frac{4}{9}\epsilon\right)^{2i}$ $\left(\frac{1}{6} - \frac{1}{18}\epsilon\right)^{2i} \left(-\frac{1}{3} + \frac{4}{9}\epsilon\right)^{n-2i}$

TABLE I. Wigner components  $w_i(n, \epsilon)$  of  $\rho_S(\epsilon)^{\otimes n}$  along with their multiplicities  $m_i(n, \epsilon)$  in decreasing order in  $i$ ,  $0 \leq i \leq n$ . The order changes depending on the noise level  $\epsilon$ , the parity of the number of copies  $n$  and the parity of the components (LHS vs RHS). Multiplication  $2i$  is considered modulo  $(n+1)$ .

where the boundary points  $(x_{-1}, L_{-1}) = (0, 0)$  and  $(x_n, L_n) = (1, 1)$  are also included. The maximum is the  $(\lfloor n/2 \rfloor)$ -th elbow and its coordinates are calculated by collecting all the positive Wigner components,

$$x_{\lfloor n/2 \rfloor} = \frac{1}{2} \left( 1 + \left( \frac{7}{9} \right)^n \right), \quad (\text{E11})$$

$$L_{\lfloor n/2 \rfloor} = \frac{1}{2} \left( 1 + \left( \frac{15 - 8\epsilon}{9} \right)^n \right). \quad (\text{E12})$$

Expressions for all the elbow coordinates follow from summing up the Wigner components in decreasing order and we present the elbow coordinates for standard Lorenz curves in Table II.

Case		$x_i$	$L_i$
$0 \leq \epsilon < \frac{3}{4}$	$n$ even	LHS: $\Phi_+(2i; n, \frac{8}{9})$ RHS: $x_{\lfloor n/2 \rfloor} + \Phi_-(2i; n, \frac{1}{9})$	$\left(\frac{5}{3} - \frac{8}{9}\epsilon\right)^n \Phi_+\left(2i; n, \frac{12-4\epsilon}{15-8\epsilon}\right)$ $L_{\lfloor n/2 \rfloor} - \left(\frac{5}{3} - \frac{8}{9}\epsilon\right)^n \Phi_-\left(2i; n, \frac{3-4\epsilon}{15-8\epsilon}\right)$
	$n$ odd	LHS: $\Phi_-(2i; n, \frac{8}{9})$ RHS: $x_{\lfloor n/2 \rfloor} + \Phi_-(2i; n, \frac{1}{9})$	$\left(\frac{5}{3} - \frac{8}{9}\epsilon\right)^n \Phi_-\left(2i; n, \frac{12-4\epsilon}{15-8\epsilon}\right)$ $L_{\lfloor n/2 \rfloor} - \left(\frac{5}{3} - \frac{8}{9}\epsilon\right)^n \Phi_-\left(2i; n, \frac{3-4\epsilon}{15-8\epsilon}\right)$
	$\frac{3}{4} \leq \epsilon < \frac{3}{2}$	$n$ even: LHS: $\Phi_+(2i; n, \frac{1}{9})$ , RHS: $x_{\lfloor n/2 \rfloor} + \Phi_-(2i; n, \frac{8}{9})$	$\left(\frac{5}{3} - \frac{8}{9}\epsilon\right)^n \Phi_+\left(2i; n, \frac{3-4\epsilon}{15-8\epsilon}\right)$ $L_{\lfloor n/2 \rfloor} - \left(\frac{5}{3} - \frac{8}{9}\epsilon\right)^n \Phi_-\left(2i; n, \frac{12-4\epsilon}{15-8\epsilon}\right)$
	$\frac{3}{2} \leq \epsilon$	$n$ odd: LHS: $\Phi_+(2i; n, \frac{1}{9})$ , RHS: $x_{\lfloor n/2 \rfloor} + \Phi_+(2i; n, \frac{8}{9})$	$\left(\frac{5}{3} - \frac{8}{9}\epsilon\right)^n \Phi_+\left(2i; n, \frac{3-4\epsilon}{15-8\epsilon}\right)$ $L_{\lfloor n/2 \rfloor} - \left(\frac{5}{3} - \frac{8}{9}\epsilon\right)^n \Phi_+\left(2i; n, \frac{12-4\epsilon}{15-8\epsilon}\right)$

TABLE II. Standard Lorenz curves elbow coordinates. The expression depends on the noise level  $\epsilon$ , the parity of the number of copies  $n$  and the location of the elbow relative to the maximum (LHS vs RHS). Multiplication  $2i$  is considered modulo  $(n+1)$ . Note that the Lorenz curve boundary points are  $(x_{-1}, L_{-1}) := (0, 0)$  and  $(x_n, L_n) = (1, 1)$ .



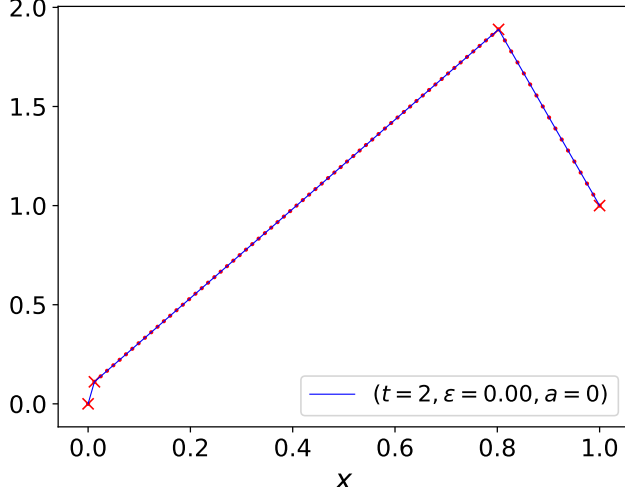


FIG. 11. **All points on the Lorenz curve are uniformly distributed.** Eqs. (E13) and (E14) capture the coordinates of all points up to the maximum.

### 3. Standard Lorenz curve coordinates

We can get explicit expressions for all  $9^n$  points of the standard Lorenz curve  $L_{\rho_S(\epsilon)^{\otimes n} \mathbb{1}/3}$ , in terms of the elbow coordinates:

$$x_{ij} = \left(1 - \frac{j}{m_i}\right) x_{i-1} + \frac{j}{m_i} x_i, \quad (\text{E13})$$

$$L_{ij} = \left(1 - \frac{j}{m_i}\right) L_{i-1} + \frac{j}{m_i} L_i \quad (\text{E14})$$

for  $j = 1, \dots, m_i$  and  $i = 0, \dots, n$ , where multiplicities  $m_i = m_i(n, \epsilon)$  are given in Table I.

For visualisation purposes, we plot the Lorenz curve of state  $\rho_S(0)^{\otimes 2}$ .

Consider the state

$$\rho_S(n', \epsilon, n - n') := \rho_S(\epsilon)^{\otimes n'} \otimes \left(\frac{1}{3} \mathbb{1}\right)^{\otimes (n - n')},$$

Tensoring with the maximally mixed state keeps the Lorenz curve unchanged, but increases the resolution of (the uniformly distributed) points. The new point coordinates are given by:

$$x_{ijk} = (1 - p_{ijk}) x_{i-1} + p_{ijk} x_i \quad (\text{E15})$$

$$L_{ijk} = (1 - p_{ijk}) L_{i-1} + p_{ijk} L_i, \quad (\text{E16})$$

$$\text{where } p_{ijk} = \frac{k + (j - 1)9^{n-n'}}{9^{n-n'} m_i}$$

for  $i = 0, \dots, n'$ ,  $j = 1, \dots, m_i(n', \epsilon')$  and  $k = 1, \dots, 9^{n-n'}$ .

We can unify the indices, by introducing a single index

$$I(i, j, k) := k + \left[(j - 1) + \sum_{\ell=0}^{i-1} m_\ell(n', \epsilon')\right] 9^{n-n'}, \quad (\text{E17})$$

so that  $I = 1, 2, \dots, 9^n$ . The elbow coordinates correspond to

$$I(i, m_i(n', \epsilon'), 9^{n-n'}) = \sum_{\ell=0}^i m_\ell(n', \epsilon'), \quad i = 0, \dots, n'. \quad (\text{E18})$$

The index function  $I$  is bijective, i.e.

$$(i, j, k) = (i', j', k') \Leftrightarrow I(i, j, k) = I(i', j', k'). \quad (\text{E19})$$

### 4. Strange state MSD in the unital fragment

Consider the Strange state MSD process in the unital fragment,

$$\rho_S(n', \epsilon, 0) \xrightarrow{\mathcal{O}_{1/3}} \rho_S(n', \epsilon', n - n'). \quad (\text{E20})$$

We denote initial state indices without a prime and target state indices with a prime,

$$I(i, j, k = 1) = j + \sum_{\ell=0}^{i-1} m_\ell(n, \epsilon), \quad (\text{E21})$$

$$I'(i', j', k') = k' + \left[(j' - 1) + \sum_{\ell=0}^{i'-1} m_\ell(n', \epsilon')\right] 9^{n-n'}. \quad (\text{E22})$$

Pointwise Lorenz curve comparison requires  $x_I = x_{I'}$ , so the question is:

*Given a triplet  $(i', j', k')$ , what is the tuple  $(i, j)$  such that  $I(i, j) = I'(i', j', k')$ ?*

According to Theorem 18 which is proved in Appendix D 4, for standard Lorenz curves, we need to match indices at the target state elbows, so the requirement on the indices is finding a tuple  $(i, j)$ , such that for a given  $i' = 0, \dots, n'$ ,

$$j + \sum_{\ell=0}^{i-1} m_\ell(n, \epsilon) = \sum_{\ell=0}^{i'-1} m_\ell(n', \epsilon'). \quad (\text{E23})$$

As a basic example, consider the process

$$\rho_S(\epsilon)^{\otimes 4} \xrightarrow{\mathcal{O}_{1/3}} \rho_S(\epsilon')^{\otimes 2} \otimes \left(\frac{1}{3} \mathbb{1}\right)^{\otimes 2}. \quad (\text{E24})$$

We want to check which Lorenz curve is higher at the first elbows of the target state, i.e. we want to verify or reject the first inequality below:

$$L_{I(i, j)} \geq L'_{I'(0, 1, 81)} \quad (\text{E25})$$

where the first multiplicity of the target state is  $m_0 = 1$ . The multiplicities of the initial state are  $(1, 384, 4096)$ .

The challenge is to find  $i, j$  such that  $i(i, j) = i'(0, 1, 81)$ . [By trial and error], we find that  $(i, j) = (1, 80)$ . Now we can use Eq. (E16) to directly calculate

$$\begin{aligned} L'_{i'(0,1,81)} &= L'_0 = \left(\frac{5}{3} - \frac{8}{9}\epsilon\right)^2 \Phi_+ \left(0; 2, 4, \frac{3-\epsilon}{15-8\epsilon}\right), \\ L_{i(1,80)} &= \left(1 - \frac{80}{384}\right) L_0 + \frac{80}{384} L_1 \\ &= \left(\frac{5}{3} - \frac{8}{9}\epsilon\right)^4 \left[ \frac{19}{24} \Phi_+ \left(0; 4, 4, \frac{3-\epsilon}{15-8\epsilon}\right) \right. \\ &\quad \left. + \frac{5}{24} \Phi_+ \left(2; 4, 4, \frac{3-\epsilon}{15-8\epsilon}\right) \right], \end{aligned}$$

and then compare them.

## Appendix F: Technical details of bound derivation in stabiliser fragments

### 1. Component-multiplicity pairs

In general, a 1-copy  $d$ -dimensional state is defined exactly by its  $d^2$  Wigner distribution  $W$ . The distribution  $W$  can always be described by a Wigner-component vector  $\mathbf{w} = (w_i)_{i=1,\dots,D}$  and a multiplicity vector  $\mathbf{m} = (m_i)_{i=1,\dots,D}$ , where  $D \leq d^2$  and they form component-multiplicity pairs denoted by  $\{(w_i, m_i)\}_{i=1,\dots,D}$ .

**Definition 21.** Consider a  $d^2$ -dimensional distribution  $W$  and a positive integer  $D$ . We call the set of order pairs  $\{(w_i, m_i)\}_{i=1,\dots,D}$  a complete set of component-multiplicity pairs, if  $W$  contains  $m_i$  components  $w_i$  and  $\sum_{i=0}^D m_i = d^2$ .

Therefore, such a set that describes a distribution  $W$  contains in its description each component of  $W$  exactly once. As an example, two complete sets of pairs for the Strange state are  $\{(-1/3, 1), (1/6, 8)\}$  and  $\{(-1/3, 1), (1/6, 2), (1/6, 3), (1/6, 3)\}$ .

Consider two states  $\rho_A, \rho_B$  with Wigner distributions  $W_{\rho_A}, W_{\rho_B}$  described respectively by complete sets of component-multiplicity pairs

$$\{(w_i, m_i)\}_{i=1,\dots,D_A} \text{ and } \{(w'_j, m'_j)\}_{j=0,\dots,D_B}. \quad (\text{F1})$$

The multiplicative property of the Wigner distribution over a composite phase space  $\mathcal{P}_{d_A} \times \mathcal{P}_{d_B}$  shown in Proposition 11,

$$W_{\rho_A \otimes \rho_B}(\mathbf{x}_A \oplus \mathbf{x}_B) = W_{\rho_A}(\mathbf{x}_A) W_{\rho_B}(\mathbf{x}_B), \quad (\text{F2})$$

implies that the distribution  $W_{\rho_A \otimes \rho_B}$  is  $d_A^2 d_B^2$ -dimensional and contains components of the form  $w_i w'_j$ . Therefore, the set  $\{(w_i w'_j, m_i m'_j)\}$  with  $i = 1, \dots, D_A$  and  $j = 1, \dots, D_B$  is a complete set of component-multiplicity pairs for the distribution of the composite system  $W_{\rho_A \otimes \rho_B}$ . This is true because all components

are of the form  $w_i w'_j$  and

$$\sum_{i=1}^{D_A} \sum_{j=1}^{D_B} m_i m'_j = \sum_{i=1}^{D_A} m_i \sum_{j=1}^{D_B} m'_j = d_A^2 d_B^2.$$

Note that the rescaled distribution is also multiplicative,

$$\begin{aligned} \widetilde{W}_{\rho_A \otimes \rho_B | \gamma_A \otimes \gamma_B}(\mathbf{x}_A \oplus \mathbf{x}_B) &= \frac{W_{\rho_A \otimes \rho_B}(\mathbf{x}_A \oplus \mathbf{x}_B)}{W_{\gamma_A \otimes \gamma_B}(\mathbf{x}_A \oplus \mathbf{x}_B)} = \\ \frac{W_{\rho_A}(\mathbf{x}_A) W_{\rho_B}(\mathbf{x}_B)}{W_{\gamma_A}(\mathbf{x}_A) W_{\gamma_B}(\mathbf{x}_B)} &= \widetilde{W}_{\rho_A | \gamma_A}(\mathbf{x}_A) \widetilde{W}_{\rho_B | \gamma_B}(\mathbf{x}_B), \quad (\text{F3}) \end{aligned}$$

so result holds for this too [FIX SECTION].

Given a state  $\rho$  and a complete set of component-multiplicity pairs describing its Wigner distribution  $W_\rho$ , we now provide a method of computing the components (and multiplicities) of the  $n$ -copy distribution  $W_\rho^{\otimes n}$ .

**Lemma 22.** Let  $W$  be a  $d$ -dimensional distribution defined by a complete set of component-multiplicity pairs  $\{(w_i, m_i)\}_{i=1,\dots,D}$  with  $D \leq d$  and consider the distribution  $W^{\otimes n}$  obtained by taking the Kronecker product  $W \otimes \dots \otimes W$  between  $n$  copies of  $W$ .

Denote by  $C_D^n := \{\mathbf{k}\}$  the set of all vectors  $\mathbf{k} := (k_1, \dots, k_D)$  with non-negative integer components that sum to  $n$ , i.e.

$$0 \leq k_1, \dots, k_D \leq n \text{ and } k_1 + \dots + k_D = n.$$

Then,  $W^{\otimes n}$  admits a complete set of component-multiplicity pairs  $\{(W_{\mathbf{k}}, M_{\mathbf{k}})\}_{\mathbf{k} \in C_D^n}$ , where

$$M_{\mathbf{k}} = \frac{n!}{k_1! \dots k_D!} \prod_{i=1}^D m_i^{k_i}, \quad (\text{F4})$$

$$W_{\mathbf{k}} = \prod_{i=1}^D w_i^{k_i}. \quad (\text{F5})$$

*Proof.* We proceed by induction.

Assume  $n = 1$ . Let  $\mathbf{k}_i := (0, \dots, 0, 1, 0, \dots, 0)$  be a vector with its  $i$ -th component equal to 1 and 0's elsewhere. The set  $C_D^1$  consists of all vectors of this form, i.e.

$$C_D^1 = \{\mathbf{k}_i\}_{i=1,\dots,D}$$

It is also true by direct calculation that

$$(W_{\mathbf{k}_i}, M_{\mathbf{k}_i}) = (w_i, m_i).$$

Therefore,  $\{(W_{\mathbf{k}}, M_{\mathbf{k}})\}_{\mathbf{k} \in C_D^1}$  is a complete set of component-multiplicity pairs for  $W$ .

Assume that  $\{(W_{\mathbf{k}}, M_{\mathbf{k}})\}_{\mathbf{k} \in C_D^n}$  as given in Eqs. (F4) and (F5) is a complete set of component-multiplicity pairs for the  $n$ -copy distribution  $W^{\otimes n}$ . By construction, the distribution  $W^{\otimes(n+1)} = W^{\otimes n} \otimes W$  is multiplicative, so it admits the complete set of component multiplicity pairs

$$\{(W_{\mathbf{k}} w_i, M_{\mathbf{k}} m_i)\}, \mathbf{k} \in C_D^n \text{ and } i = 1, \dots, D. \quad (\text{F6})$$

Consider the component sum of the distribution  $W^{\otimes(n+1)}$ ,

$$\begin{aligned} \sum_{\mathbf{k} \in C_D^n} \sum_{i=1}^D M_{\mathbf{k}} m_i W_{\mathbf{k}} w_i &= \sum_{\mathbf{k} \in C_D^n} M_{\mathbf{k}} W_{\mathbf{k}} \sum_{i=1}^D m_i w_i = \\ \sum_{\mathbf{k} \in C_D^n} \frac{n!}{k_1! \dots k_D!} \prod_{i=1}^D m_i^{k_i} w_i^{k_i} \sum_{i=1}^D m_i w_i &= \\ \left( \sum_{i=1}^D m_i w_i \right)^n \left( \sum_{i=1}^D m_i w_i \right) &= \left( \sum_{i=1}^D m_i w_i \right)^{n+1} = \\ \sum_{\mathbf{q} \in C_D^{n+1}} M_{\mathbf{q}} W_{\mathbf{q}}, \end{aligned}$$

where in the last expression, vectors  $\mathbf{q} = (q_1, \dots, q_D)$  have non-negative integer components that sum to  $(n+1)$  and

$$\begin{aligned} M_{\mathbf{q}} &= \frac{(n+1)!}{q_1! \dots q_D!} \prod_{i=1}^D m_i^{q_i}, \\ W_{\mathbf{q}} &= \prod_{i=1}^D w_i^{q_i}. \end{aligned}$$

We have used the multinomial theorem to proceed between lines 2-3 and lines 3-4 of the derivation.

We have achieved a regrouping of the distribution components. Every component  $W_{\mathbf{q}}$  is of the form  $W_{\mathbf{k}} w_i$  with  $q_i = k_i + 1$  and  $q_j = k_j$  for  $j \neq i$  and

$$\sum_{\mathbf{q} \in C_D^{n+1}} M_{\mathbf{q}} = \sum_{\mathbf{q} \in C_D^{n+1}} \frac{(n+1)!}{q_1! \dots q_D!} \prod_{i=1}^D m_i^{q_i} = \left( \sum_{i=1}^D m_i \right)^{n+1} = d^{n+1},$$

which is the dimension of  $W^{\otimes(n+1)}$ .

Therefore,  $\{(W_{\mathbf{q}}, M_{\mathbf{q}})\}_{\mathbf{q} \in C_D^{n+1}}$  is a complete set of component-multiplicity pairs for  $W^{\otimes n}$ , completing the proof.  $\square$

## 2. First elbow location

Here, we show that the horizontal location of the first elbow is lower for an initial state compared to the target state for a valid distillation process ( $n > n'$ ).

1. (C1)  $\rightarrow$  (C1). It is always true that  $e^{-\beta E_0} / \mathcal{Z}_{\beta} \leq 1$ , so

$$\begin{aligned} x_0 &= \left( \frac{e^{-\beta E_0}}{3\mathcal{Z}_{\beta}} \right)^n = \left( \frac{e^{-\beta E_0}}{3\mathcal{Z}_{\beta}} \right)^{n-n'} \left( \frac{e^{-\beta E_0}}{3\mathcal{Z}_{\beta}} \right)^{n'} \\ &< \left( \frac{e^{-\beta E_0}}{3\mathcal{Z}_{\beta}} \right)^{n'} = x'_0 \end{aligned}$$

2. (C2)  $\rightarrow$  (C1). It is always true that

$$\begin{aligned} \mathcal{Z}_{\beta}^{n-n'} &> (e^{-\beta E_{\max}})^{n-n'} = e^{-\beta(n-n')E_{\max}}, \text{ so} \\ x_0 &= \left( \frac{e^{-\beta E_{\max}}}{\mathcal{Z}_{\beta}} \right)^n = \frac{e^{-\beta(nE_{\max}-n'E_0)}}{\mathcal{Z}_{\beta}^{n-n'}} \left( \frac{e^{-\beta E_0}}{3\mathcal{Z}_{\beta}} \right)^{n'} \\ &< \frac{e^{-\beta(nE_{\max}-n'E_0)}}{e^{-\beta(n-n')E_{\max}}} \left( \frac{e^{-\beta E_0}}{3\mathcal{Z}_{\beta}} \right)^{n'} \\ &= e^{-\beta n'(E_{\max}-E_0)} \left( \frac{e^{-\beta E_0}}{3\mathcal{Z}_{\beta}} \right)^{n'} \\ &\leq \left( \frac{e^{-\beta E_0}}{3\mathcal{Z}_{\beta}} \right)^{n'} = x'_0. \end{aligned}$$

3. (C2)  $\rightarrow$  (C2). It is always true that  $e^{-\beta E_{\max}} / \mathcal{Z}_{\beta} \leq 1$ , so

$$\begin{aligned} x_0 &= \left( \frac{e^{-\beta E_{\max}}}{\mathcal{Z}_{\beta}} \right)^n = \left( \frac{e^{-\beta E_{\max}}}{\mathcal{Z}_{\beta}} \right)^{n-n'} \left( \frac{e^{-\beta E_{\max}}}{\mathcal{Z}_{\beta}} \right)^{n'} \\ &\leq \left( \frac{e^{-\beta E_{\max}}}{\mathcal{Z}_{\beta}} \right)^{n'} = x'_0 \end{aligned}$$