

## SUPPLEMENTARY NOTE 1: PROPERTIES OF WIGNER REPRESENTATIONS

Here we present basic properties of the phase-point operators and the Wigner distribution that are used throughout the main text.

**Proposition 1.** *For any dimension  $d$ , the phase-point operators satisfy:*

- (i) *Hermiticity and unitarity:*  $A_{\mathbf{z}}^\dagger = A_{\mathbf{z}} = A_{\mathbf{z}}^{-1}$ ;
- (ii) *Closure under transposition:*  $A_{(q,p)}^T = A_{(q,-p)}$ ;
- (iii) *Unit trace for odd  $d$ :*  $\text{tr}[A_{\mathbf{z}}] = 1$ ;
- (iv) *Completeness relation:*  $\sum_{\mathbf{z} \in \mathcal{P}_d} A_{\mathbf{z}} = d\mathbb{1}$ ;
- (i) *Orthogonality:*  $\text{tr}[A_{\mathbf{z}}^\dagger A_{\mathbf{y}}] = d\delta_{\mathbf{z},\mathbf{y}}$ .

All properties follow from the definition in Eq. (6) in the main text along with properties of the displacement operators  $D_{\mathbf{z}}$  and can be found in the literature, e.g. [1–3]

**Proposition 2.** *The state Wigner distribution is*

- (i) *Real valued:*  $W_\rho(\mathbf{z}) \in \mathbb{R}^{d^2}$ ;
- (ii) *Normalized:*  $\sum_{\mathbf{z} \in \mathcal{P}_d} W_\rho(\mathbf{z}) = 1$ ;
- (iii) *Bounded:*  $|W_\rho(\mathbf{z})| \leq \frac{1}{d}$ .
- (iv) *Additive under mixing:*  
 $W_{p\rho_1 + (1-p)\rho_2}(\mathbf{z}) = pW_{\rho_1}(\mathbf{z}) + (1-p)W_{\rho_2}(\mathbf{z})$ ;
- (v) *Multiplicative under tensor products:*  
 $W_{\rho_A \otimes \rho_B}(\mathbf{z}_A \oplus \mathbf{z}_B) = W_{\rho_A}(\mathbf{z}_A)W_{\rho_B}(\mathbf{z}_B)$ .

*Proof.* Proof of all properties can be found in the literature [1–4] except for property (iii) which we prove here.

Let  $\{\lambda_i\}_{i \in \mathbb{Z}_d}$  be the (non-negative) eigenvalues of  $\rho$ , summing to 1. Let  $\{\alpha_{\mathbf{z},i}\}_{i \in \mathbb{Z}_d}$  be the eigenvalues of  $A_{\mathbf{z}}$ . For any  $\mathbf{z}, \alpha_{\mathbf{z},i} \in \{-1, 1\}$ , due to the hermiticity and unitarity of the phase-point operators. Then,

$$\begin{aligned} |W_\rho(\mathbf{z})| &= \frac{1}{d} |\text{tr}[A_{\mathbf{z}}\rho]| \leq \frac{1}{d} \left| \sum_i \alpha_{\mathbf{z},i} \lambda_i \right| \\ &\leq \frac{1}{d} \sum_i \lambda_i = \frac{1}{d}. \end{aligned} \quad (1)$$

The first inequality follows from Theorem 1 of [5] for the trace of complex matrices, while the second is the Cauchy-Schwarz inequality.  $\square$

**Proposition 3.** *The Wigner distribution of a quantum channel  $\mathcal{E} : \mathcal{B}(\mathcal{H}_{d_A}) \rightarrow \mathcal{B}(\mathcal{H}_{d_B})$  is*

- (i) *Real-valued:*  $W_{\mathcal{E}}(\mathbf{z}|\mathbf{y}) \in \mathbb{R}^{d_A^2} \times \mathbb{R}^{d_B^2}$ ;
- (ii) *Normalized:*  $\sum_{\mathbf{y} \in \mathcal{P}_{d_B}} W_{\mathcal{E}}(\mathbf{z}|\mathbf{y}) = 1$   
for any  $\mathbf{z} \in \mathcal{P}_{d_A}$ ;

(iii) *Bounded:*  $|W_{\mathcal{E}}(\mathbf{z}|\mathbf{y})| \leq \frac{d_A}{d_B}$ ;

(iv)  $W_{\mathcal{E}(\rho)}(\mathbf{z}) = \sum_{\mathbf{y} \in \mathcal{P}_{d_A}} W_{\mathcal{E}}(\mathbf{z}|\mathbf{y})W_\rho(\mathbf{y})$  for any  $\mathbf{z} \in \mathcal{P}_{d_B}$ .

*Proof.* Proof of all properties are provided in [4] except for property (iii) which is a direct consequence of the definition of  $W_{\mathcal{E}}$  and property (iii) in Proposition 2.  $\square$

Within the context of the Wigner representation, we define the rescaled quasi-distribution

$$W_{\rho|\tau}(\mathbf{z}) := \frac{W_\rho(\mathbf{z})}{W_\tau(\mathbf{z})}, \quad (2)$$

which is well-defined if  $\tau$  is a full-rank stabilizer state.

The following result shows that the rescaled Lorenz curves behave in a natural way under tensor products.

**Proposition 4.** *Let  $\tau_A, \tau_B$  be full rank stabilizer states on systems  $A$  and  $B$ , and let  $\rho_A, \rho_B$  be arbitrary states on  $A, B$ . Then, the rescaled quasi-distribution obeys*

$$W_{\rho_A \otimes \rho_B | \tau_A \otimes \tau_B}(\mathbf{z}_A \oplus \mathbf{z}_B) = W_{\rho_A | \tau_A}(\mathbf{z}_A)W_{\rho_B | \tau_B}(\mathbf{z}_B) \quad (3)$$

*Proof.* This follows from the multiplicativity of the Wigner distribution,

$$\begin{aligned} W_{\rho_A \otimes \rho_B | \tau_A \otimes \tau_B}(\mathbf{z}_A \oplus \mathbf{z}_B) &= \frac{W_{\rho_A \otimes \rho_B}(\mathbf{z}_A \oplus \mathbf{z}_B)}{W_{\tau_A \otimes \tau_B}(\mathbf{z}_A \oplus \mathbf{z}_B)} \\ &= \frac{W_{\rho_A}(\mathbf{z}_A)W_{\rho_B}(\mathbf{z}_B)}{W_{\tau_A}(\mathbf{z}_A)W_{\tau_B}(\mathbf{z}_B)} = W_{\rho_A | \tau_A}(\mathbf{z}_A)W_{\rho_B | \tau_B}(\mathbf{z}_B). \end{aligned}$$

$\square$

## SUPPLEMENTARY NOTE 2: PROPERTIES OF RELATIVE MAJORIZATION AND LORENZ CURVES

### Relative majorization for quasi-distributions

We prove the following result connecting majorization to relative majorization [6–8].

**Proposition 5.** *Let  $\mathbf{w} \in \mathbb{R}^N, \mathbf{w}' \in \mathbb{R}^{N'}$  be quasi-distributions and  $\mathbf{r} \in \mathbb{R}^N, \mathbf{r}' \in \mathbb{R}^{N'}$  probability distributions with positive rational entries given by  $r_i = a_i/K$  and  $r'_i = a'_i/K$  for positive integers  $a_i, a'_i$  and  $K = \sum_{i=1}^N a_i = \sum_{i=1}^{N'} a'_i$ . Then,*

$$(\mathbf{w}, \mathbf{r}) \succ (\mathbf{w}', \mathbf{r}') \text{ if and only if } \Gamma_{\mathbf{a}}(\mathbf{w}) \succ \Gamma_{\mathbf{a}'}(\mathbf{w}'), \quad (4)$$

where the embedding map  $\Gamma_{\mathbf{a}} : \mathbb{R}^N \rightarrow \mathbb{R}^K$  is given by

$$\Gamma_{\mathbf{a}}(\mathbf{z}) := \bigoplus_{i=1}^N z_i \boldsymbol{\eta}_{a_i}, \quad (5)$$

with  $\boldsymbol{\eta}_{a_i} = (1/a_i, 1/a_i, \dots, 1/a_i)$  the uniform distribution on  $a_i$  elements.

*Proof.* We first note that the map  $\Gamma_{\mathbf{a}}$  is stochastic and has a well-defined left-inverse  $\Gamma_{\mathbf{a}}^{-1} : \mathbb{R}^K \rightarrow \mathbb{R}^N$  given by

$$\Gamma_{\mathbf{a}}^{-1}(\mathbf{x}) = \left( \sum_{i=1}^{a_1} x_i, \sum_{i=a_1+1}^{a_1+a_2} x_i, \dots, \sum_{i=a_1+\dots+a_{N-1}+1}^K x_i \right), \quad (6)$$

that obeys  $(\Gamma_{\mathbf{a}}^{-1} \circ \Gamma_{\mathbf{a}})(\mathbf{z}) = \mathbf{z}$  for all  $\mathbf{z} \in \mathbb{R}^N$ .

The claim is equivalent to the statement that there exists a bistochastic map  $B$  sending  $\Gamma_{\mathbf{a}}(\mathbf{w})$  to  $\Gamma_{\mathbf{a}'}(\mathbf{w}')$  if and only if there exists a stochastic map  $A$  sending  $\mathbf{w}$  to  $\mathbf{w}'$  and  $\mathbf{r}$  to  $\mathbf{r}'$ .

Suppose there is a stochastic map  $A$  such that  $A\mathbf{w} = \mathbf{w}'$  and  $A\mathbf{r} = \mathbf{r}'$ . We define  $B : \mathbb{R}^K \mapsto \mathbb{R}^K$  by  $B := \Gamma_{\mathbf{a}'} \circ A \circ \Gamma_{\mathbf{a}}^{-1}$  so that it is stochastic as a composition of stochastic maps and it preserves the uniform distribution in  $\mathbb{R}^K$ , since

$$\begin{aligned} B(1/K, 1/K, \dots, 1/K) &= (\Gamma_{\mathbf{a}'} \circ A \circ \Gamma_{\mathbf{a}}^{-1})(\Gamma_{\mathbf{a}}(\mathbf{r})) = \\ &= \Gamma_{\mathbf{a}'}(\mathbf{r}') = (1/K, 1/K, \dots, 1/K), \end{aligned} \quad (7)$$

therefore  $B$  is bistochastic. Finally,  $B$  maps the embedded distributions as follows,

$$B\Gamma_{\mathbf{a}}(\mathbf{w}) = (\Gamma_{\mathbf{a}'} \circ A \circ \Gamma_{\mathbf{a}}^{-1})(\Gamma_{\mathbf{a}}(\mathbf{w})) = \Gamma_{\mathbf{a}'}(\mathbf{w}'). \quad (8)$$

Conversely, suppose a bistochastic map  $B$  exists sending  $\Gamma_{\mathbf{a}}(\mathbf{w})$  to  $\Gamma_{\mathbf{a}'}(\mathbf{w}')$ . Again, define  $A : \mathbb{R}^N \mapsto \mathbb{R}^{N'}$  by  $A := \Gamma_{\mathbf{a}'}^{-1} \circ B \circ \Gamma_{\mathbf{a}}$  so that it is stochastic as a composition of stochastic maps and  $A\mathbf{w} = (\Gamma_{\mathbf{a}'}^{-1} \circ B \circ \Gamma_{\mathbf{a}})\mathbf{w} = \mathbf{w}'$ , as well as  $A\mathbf{r} = \mathbf{r}'$ .  $\square$

We note that the vectors  $\mathbf{r}, \mathbf{r}'$  with rational components form a dense subset of the positive probability distributions, and do not consider further technicalities that have no impact on actual physical measurements, which always have a finite resolution.

Restricting the above result in the context of proper probability distributions, we can prove the Lorenz curve condition for probability distributions. We define a Lorenz curve elbow as a point where the slope of the curve changes, expressed explicitly in Eq. (15) for Lorenz curve  $L_{\mathbf{w}|\mathbf{r}}(x)$

**Proposition 6.** *Given probability distributions  $\mathbf{w}, \mathbf{r} \in \mathbb{R}^N$  and  $\mathbf{w}', \mathbf{r}' \in \mathbb{R}^{N'}$  with  $\mathbf{r}, \mathbf{r}'$  having positive components, then*

$$(\mathbf{w}, \mathbf{r}) \succ (\mathbf{w}', \mathbf{r}') \text{ if and only if } L_{\mathbf{w}|\mathbf{r}}(x) \geq L_{\mathbf{w}'|\mathbf{r}'}(x),$$

for all  $x \in [0, 1]$ .

*Proof.* In the case of  $N = N'$  and  $\mathbf{r} = \mathbf{r}' = \boldsymbol{\eta}$ , where  $\boldsymbol{\eta} = (1/N, 1/N, \dots, 1/N)$  is the uniform distribution on  $N$  elements, the statement reduces to the Lorenz curve condition  $L_{\mathbf{w}}(x) \geq L_{\mathbf{w}'}(x)$  for all  $x$ , for simple majorization which follows immediately from the defining set of inequalities for majorization [9]. Namely, the Lorenz curve  $L_{\mathbf{w}}(x)$  for  $\mathbf{w}$  is obtained from the partial sums of  $\mathbf{w}$  sorted in non-increasing order. It is also clear from

the definition that the Lorenz curve of  $\mathbf{w}$  is given by  $L_{\mathbf{w}}(x) = L_{\mathbf{w}|\boldsymbol{\eta}}(x)$ .

To prove the general statement we reduce the relative majorization Lorenz curve condition to standard majorization. Using the notation and assumptions of Proposition 5 for distributions with rational components, we can define  $\Gamma_{\mathbf{a}}(\mathbf{w}), \Gamma_{\mathbf{a}'}(\mathbf{w}')$  and the uniform distribution  $\boldsymbol{\eta} = (1/K, 1/K, \dots, 1/K)$ .

The key ingredient in the proof is that the Lorenz curve of  $\mathbf{w}$  relative to  $\mathbf{r}$  coincides with the Lorenz curve of  $\Gamma_{\mathbf{a}}(\mathbf{w})$ , namely

$$L_{\mathbf{w}|\mathbf{r}}(x) = L_{\Gamma_{\mathbf{a}}(\mathbf{w})|\boldsymbol{\eta}}(x) = L_{\Gamma_{\mathbf{a}}(\mathbf{w})}(x) \text{ for all } x \in [0, 1]. \quad (9)$$

To see this we consider the elbows of  $L_{\mathbf{w}|\mathbf{r}}(x)$ ,

$$(x_k, L_{\mathbf{w}|\mathbf{r}}(x_k)) = \left( \sum_{i=1}^k r_{\pi(i)}, \sum_{i=1}^k w_{\pi(i)} \right), \quad (10)$$

where the permutation  $\pi$  sorts  $(w_i/r_i)$  in non-increasing order. Expressing  $\Gamma_{\mathbf{a}}(\mathbf{w})$  as

$$\Gamma_{\mathbf{a}}(\mathbf{w}) = \frac{1}{K} \bigoplus_{i=1}^N \left( \frac{w_i}{r_i}, \dots, \frac{w_i}{r_i} \right), \quad (11)$$

where  $(w_i/r_i, \dots, w_i/r_i)$  has  $a_i$  elements, it is clear that permutation  $\pi$  sorts  $\Gamma_{\mathbf{a}}(\mathbf{w})$  in non-increasing order too. The Lorenz curve elbows  $(y_k, L_{\Gamma_{\mathbf{a}}(\mathbf{w})|\boldsymbol{\eta}}(y_k))$  occur at

$$y_k = \sum_{j=1}^{a_{\pi(1)} + \dots + a_{\pi(k)}} \frac{1}{K} = \sum_{i=1}^k \frac{a_{\pi(i)}}{K} = x_k \quad (12)$$

and take values

$$L_{\Gamma_{\mathbf{a}}(\mathbf{w})|\boldsymbol{\eta}}(y_k) = \sum_{i=1}^k a_{\pi(i)} \frac{1}{K} \frac{w_{\pi(i)}}{r_{\pi(i)}} = L_{\mathbf{w}|\mathbf{r}}(x_k). \quad (13)$$

Therefore, the Lorenz curve of the embedded distribution coincides with the Lorenz curve in the relative majorization setting.

Finally, we have  $(\mathbf{w}, \mathbf{r}) \succ (\mathbf{w}', \mathbf{r}')$  if and only if  $\Gamma_{\mathbf{a}}(\mathbf{w}) \succ \Gamma_{\mathbf{a}'}(\mathbf{w}')$ , which holds if and only if  $L_{\Gamma_{\mathbf{a}}(\mathbf{w})|\boldsymbol{\eta}}(x) \geq L_{\Gamma_{\mathbf{a}'}(\mathbf{w}')|\boldsymbol{\eta}}(x)$  for all  $x \in [0, 1]$ , which in turn holds if and only if  $L_{\mathbf{w}|\mathbf{r}}(x) \geq L_{\mathbf{w}'|\mathbf{r}'}(x)$ ,  $x \in [0, 1]$ , which concludes the proof.  $\square$

We now verify useful majorization properties in the context of quasi-probability distributions, starting from the property that the Lorenz curve is concave.

**Proposition 7.** *Let  $\mathbf{w}$  be a quasi-distribution and let  $\mathbf{r}$  be a probability distribution with strictly non-zero components. Then  $L_{\mathbf{w}|\mathbf{r}}(x)$  is a concave function on  $[0, 1]$ .*

*Proof.* Let  $x_*$  be the point where  $L_{\mathbf{w}|\mathbf{r}}(x)$  first attains its maximum. Therefore, on  $[0, x_*]$  the function rises monotonically to  $L_{\mathbf{w}|\mathbf{r}}(x_*)$ , via the sum of all positive entries of  $\mathbf{w}$ , taken in decreasing order. Likewise on  $[x_*, 1]$ , the

function decreases monotonically from its maximum via the partial sums of the negative entries of  $\mathbf{w}$  in decreasing order. Let  $f_1(x)$  be equal to  $L_{\mathbf{w}|\mathbf{r}}(x)$  on  $[0, x_*)$  and 0 otherwise. Also let  $f_2(x)$  be equal to  $L_{\mathbf{w}|\mathbf{r}}(x)$  on  $[x_*, 1]$  and 0 otherwise. By inspection both  $f_1$  and  $f_2$  are concave functions, and  $f_1(x) + f_2(x) = L_{\mathbf{w}|\mathbf{r}}(x)$  for all  $x \in [0, 1]$ . However, the sum of two concave functions is also concave which concludes the proof.  $\square$

The following result is used in the main text to extend the Lorenz curve condition of Proposition 6 into the context of quasi-distributions.

**Proposition 8.** *Let  $\mathbf{w}$  be a quasi-distribution and let  $\mathbf{r}$  be a probability distribution with strictly non-zero components. Then,  $L_{a\mathbf{w}+b\mathbf{r}}(x) = aL_{\mathbf{w}|\mathbf{r}}(x) + bx$  for any constants  $a > 0$  and  $b \in \mathbb{R}$ .*

*Proof.* The Lorenz curve of  $a\mathbf{w} + b\mathbf{r}$  relative to  $\mathbf{r}$  passes through  $(0, 0)$  and the points  $(\sum_{i=1}^k r_{\pi(i)}, \sum_{i=1}^k (a\mathbf{w} + b\mathbf{r})_{\pi(i)})$  where  $\pi$  is the permutation that puts  $(aw_i/r_i + b)$  in non-increasing order. Since  $a > 0$ , the permutation  $\pi$  puts  $(w_i/r_i)$  in non-increasing order too. We thus have

$$\left( \sum_{i=1}^k r_{\pi(i)}, \sum_{i=1}^k (a\mathbf{w} + b\mathbf{r})_{\pi(i)} \right) = \left( \sum_{i=1}^k r_{\pi(i)}, a \sum_{i=1}^k w_{\pi(i)} + b \sum_{i=1}^k r_{\pi(i)} \right),$$

so the value of the Lorenz curve at each potential elbow point  $x_k = \sum_{i=1}^k r_{\pi(i)}$  is given by

$$\begin{aligned} L_{a\mathbf{w}+b\mathbf{r}}(x_k) &= aL_{\mathbf{w}|\mathbf{r}}(x_k) + bL_{\mathbf{r}}(x_k) = \\ &= aL_{\mathbf{w}|\mathbf{r}}(x_k) + bx_k, \end{aligned} \quad (14)$$

so we have  $L_{a\mathbf{w}+b\mathbf{r}}(x) = aL_{\mathbf{w}|\mathbf{r}}(x) + bx$  for any  $x \in [0, 1]$  due to linearity.  $\square$

The following provides equivalent formulations of relative majorization on quasi-distributions.

**Proposition 9.** *Given quasi-distributions  $\mathbf{w}, \mathbf{w}'$ ,  $\mathbf{r}, \mathbf{r}'$ , such that the components of  $\mathbf{r}$  and  $\mathbf{r}'$  are positive, the following statements are equivalent:*

(i)  $\mathbf{w}' = A\mathbf{w}$  and  $\mathbf{r}' = A\mathbf{r}$  for a stochastic map  $A$ ;

(ii)  $L_{\mathbf{w}|\mathbf{r}}(t) \geq L_{\mathbf{w}'|\mathbf{r}'}(t)$  for  $t \in [0, 1]$  and

$$L_{\mathbf{w}|\mathbf{r}}(1) = L_{\mathbf{w}'|\mathbf{r}'}(1);$$

(iii)  $\sum_{i=1}^n |w_i - r_i t| \geq \sum_{i=1}^n |w'_i - r'_i t|$  for all  $t \in \mathbb{R}$ .

*Proof.* The proofs for these properties on proper probability distributions can be found in [9–12] and references therein.

The equivalence between (i) and (ii) on quasi-distributions was proven in Theorem 1 in the main text.

The equivalence between (i) and (iii) follows from a similar argument where we mask negativity with a probability distribution. Namely, let  $\epsilon > 0$  be such that  $\mathbf{w}(\epsilon) := \epsilon\mathbf{w} + (1 - \epsilon)\mathbf{r}$  and  $\mathbf{w}'(\epsilon) := \epsilon\mathbf{w}' + (1 - \epsilon)\mathbf{r}'$  are genuine probability distributions. This is guaranteed by picking a sufficiently small  $\epsilon > 0$ , since  $\mathbf{r}$  has positive components. We now have that  $(\mathbf{w}, \mathbf{r}) \succ (\mathbf{w}', \mathbf{r}')$  if and only if  $(\mathbf{w}(\epsilon), \mathbf{r}) \succ (\mathbf{w}'(\epsilon), \mathbf{r}')$ . Moreover, we have that for all  $c \in \mathbb{R}$ ,  $\sum_{i=1}^n |w(\epsilon)_i - r_i c| \geq \sum_{i=1}^n |w'(\epsilon)_i - r'_i c|$ , which leads to

$$\sum_{i=1}^n \left| w_i - r_i \frac{\epsilon + c - 1}{\epsilon} \right| \geq \sum_{i=1}^n \left| w'_i - r'_i \frac{\epsilon + c - 1}{\epsilon} \right|$$

Replacing  $t = (\epsilon + c - 1)/\epsilon$ , we see that  $t$  can attain any real value for  $c \in \mathbb{R}$ , so we deduce the required  $L_1$ -norm condition on quasi-distributions  $\mathbf{w}, \mathbf{w}'$ .  $\square$

### Component-multiplicity pairs

In general, a 1-copy  $d$ -dimensional state  $\rho$  is described by its  $d^2$ -dimensional Wigner distribution  $W_\rho$ . The distribution  $W_\rho$  is defined on the phase space, but it can be convenient to re-express this using vector notation. We discuss this in terms of Wigner distributions, but there is nothing to prevent the discussion from applying to rescaled Wigner distributions as well.

To each Wigner distribution  $W_\rho(\mathbf{z})$  we can associate component-multiplicity pairs  $\{(w_i, m_i)\}$  where the value  $w_i$  occurs in the distribution  $W_\rho(\mathbf{z})$  with multiplicity  $m_i$ .

As an example, for the Strange state  $\rho_S := \rho_S(\epsilon = 0)$  with

$$W_{\rho_S} = (-1/3, 1/6, 1/6, 1/6, 1/6, 1/6, 1/6, 1/6, 1/6) \quad (15)$$

we have the component-multiplicity pairs:  $\{(-1/3, 1), (1/6, 8)\}$ . However, we might also wish more freedom and not require that the different  $w_i$  values are all distinct. For example, the component-multiplicity pairs  $\{(-1/3, 1), (1/6, 2), (1/6, 3), (1/6, 3)\}$  also describe  $W_{\rho_S}$ .

This representation is more compact when a Wigner distribution has a lot of multiplicities, and allows for simple handling of multiple copies via the following fact.

**Proposition 10.** *Consider two Wigner distributions  $W_{\rho_A}(\mathbf{z}_A), W_{\rho_B}(\mathbf{z}_B)$  with component-multiplicity pairs*

$$\{(w_i, m_i)\} \text{ and } \{(w'_j, m'_j)\}, \quad (16)$$

*respectively. Then,  $\{(w_i w'_j, m_i m'_j)\}$  gives component-multiplicity pairs for  $W_{\rho_A \otimes \rho_B}(\mathbf{z}_A \oplus \mathbf{z}_B)$ .*

*Proof.* This result is true because all components of  $W_{\rho_A \otimes \rho_B}(\mathbf{z}_A \oplus \mathbf{z}_B)$  are of the form  $w_i w'_j$  and

$$\sum_i \sum_j m_i m'_j = \sum_i m_i \sum_j m'_j = d_A^2 d_B^2,$$

where  $d_A, d_B$  are the dimensions of  $\rho_A, \rho_B$  respectively, and so the set  $\{(w_i w'_j, m_i m'_j)\}$  contains exactly the Wigner components of  $W_{\rho_A \otimes \rho_B}(\mathbf{z}_A \oplus \mathbf{z}_B)$ .  $\square$

The above result also applies when we consider rescaled Wigner distributions, crucially due to the multiplicativity result shown in Proposition 4.

In the case of distillation protocols, we are interested in copies of distributions. For this, we have the following result, which follows from combinatorics.

**Proposition 11.** *Suppose  $W_\rho$  has a set of  $D$  component-multiplicity pairs  $\{(w_i, m_i)\}$ . Then,  $W_{\rho^{\otimes n}}$  has component-multiplicity pairs  $\{(W_{\mathbf{q}}, M_{\mathbf{q}})\}$ , with index  $\mathbf{q}$  running through all vectors  $(q_1, \dots, q_D)$ , where  $q_1, \dots, q_D$  are non-negative integers that sum to  $n$ , and*

$$W_{\mathbf{q}} = \prod_{i=1}^D w_i^{q_i} \quad (17)$$

$$M_{\mathbf{q}} = \binom{n}{q_1, q_2, \dots, q_D} \prod_{i=1}^D m_i^{q_i}. \quad (18)$$

The term outside the product in the expression for  $M_{\mathbf{q}}$  is the generalized binomial coefficient,

$$\binom{n}{q_1, q_2, \dots, q_D} := \frac{n!}{q_1! \dots q_D!}. \quad (19)$$

*Proof.* 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.

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

We proceed by induction. Assume  $n = 1$  and let  $\mathbf{k}_i$  be the 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  $W_\rho$ .

Assume that  $\{(W_{\mathbf{k}}, M_{\mathbf{k}})\}_{\mathbf{k} \in C_D^n}$  as given in Eqs. (17, 18) is a complete set of component-multiplicity pairs for the  $n$ -copy distribution  $W_{\rho^{\otimes n}} = W_\rho^{\otimes n}$ . By construction, the distribution  $W_{\rho^{\otimes(n+1)}} = W_\rho^{\otimes n} \otimes W_\rho$ , so it admits the complete set of component multiplicity pairs

$$\{(W_{\mathbf{k}} w_i, M_{\mathbf{k}} m_i)\}, \mathbf{k} \in C_D^n \text{ and } i = 1, \dots, D. \quad (20)$$

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

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

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

$$\begin{aligned} \sum_{\mathbf{q} \in C_D^{n+1}} M_{\mathbf{q}} &= \sum_{\mathbf{q} \in C_D^{n+1}} \frac{(n+1)!}{q_1! \dots q_D!} \prod_{i=1}^D m_i^{q_i} = \left( \sum_{i=1}^D m_i \right)^{n+1} \\ &= d^{2(n+1)}, \end{aligned}$$

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

Therefore,  $\{(W_{\mathbf{q}}, M_{\mathbf{q}})\}_{\mathbf{q} \in C_D^{n+1}}$  contains exactly the components of  $W_\rho^{\otimes(n+1)}$ , completing the proof.  $\square$

Once again, the above result applies on rescaled Wigner distributions as well.

## Results on Lorenz curve constraints

We first prove the following result on the relation between majorization in a particular sub-theory  $\mathcal{R}_\sigma$  and mana.

**Theorem 5.** *Given a magic state  $\rho$ , the maximum  $L_\star$  of its Lorenz curve  $L_{\rho|\sigma}(x)$  is independent of the  $\mathcal{R}_\sigma$  and equal to  $1 + \text{sn}(\rho)$ . Moreover, the majorization constraint is stronger than mana in every  $\mathcal{R}_\sigma$ .*

*Proof.* We denote the Wigner distributions of the states single-component vectors  $\mathbf{w}(\rho) = W_\rho(\mathbf{z})$  and  $\mathbf{w}(\sigma) = W_\sigma(\mathbf{z})$ . Likewise, we write  $\mathbf{w}(\rho|\sigma) = W_{\rho|\tau}(\mathbf{z})$ . We choose

the component indexing so that  $\mathbf{w}(\rho|\sigma)^\downarrow = \mathbf{w}(\rho|\sigma)$ , so the components are sorted in non-increasing order.

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

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

$$L_{\rho|\sigma}(x_{i_\star}) = \sum_{i=1}^{i_\star} w(\rho)_i, \quad (21)$$

and is achieved at  $x_{i_\star} := \sum_{i=1}^{i_\star} w(\sigma)_i$ . The location of this maximum varies with  $\mathcal{R}_\sigma$ , but its value is independent of  $\sigma$ ,

$$\begin{aligned} L_\star &:= L_{\rho|\sigma}(x_{i_\star}) = \sum_{\mathbf{z}: W_\rho(\mathbf{z}) \geq 0} W_\rho(\mathbf{z}) \\ &= 1 + \text{sn}(\rho). \end{aligned} \quad (22)$$

Since mana is a monotonic function of sum-negativity,  $\mathcal{M}(\rho) = \log(2\text{sn}(\rho)+1)$ , we see that mana determines the peak of the Lorenz curve  $L_{\rho|\sigma}(x)$ . However, mana is one of  $d^{2n}$  constraints, so majorization is strictly a stronger constraint in any  $\mathcal{R}_\sigma$ .  $\square$

We now prove a simple majorization constraint, arising by considering only the part of the Lorenz curves between the origin  $(0,0)$  and the first elbow.

**Proposition 12** (First elbow constraint). *Consider a magic state process  $\rho \rightarrow \rho'$  with input and output Lorenz curves  $L_{\rho|\sigma}(x), L_{\rho'|\sigma'}(x)$  and denote the coordinates of the first elbow of  $L_{\rho|\sigma}(x)$  by  $(X_0, L_0)$  and the coordinates of the first elbow of  $L_{\rho'|\sigma'}(x)$  by  $(X'_0, L'_0)$ .*

*Then, given any coordinates  $(x, L)$  and  $(x', L')$  on the input and output Lorenz curves respectively, where  $0 < x \leq X_0$  and  $0 < x' \leq X'_0$ , the process is possible only if*

$$\frac{L}{x} \geq \frac{L'}{x'}. \quad (23)$$

*Proof.* If the transformation is possible, then  $L_{\rho|\sigma}(x) \geq L_{\rho'|\sigma'}(x)$  for all  $x$  in  $[0, 1]$ . Restricting to the initial linear segment of  $L_{\rho|\sigma}(x)$  joining  $(0,0)$  to  $(X_0, L_0)$ , we then require that it be above the initial linear segment of  $L_{\rho'|\sigma'}(x)$  joining  $(0,0)$  to  $(X'_0, L'_0)$ . Since both these linear segments start at the origin, this condition is equivalent to the slope of the initial segment of the input Lorenz curve being greater than or equal to the slope of the initial segment of the output Lorenz curve. This slope constraint is then equivalent to Eq. (23).  $\square$

The following result reduces the Lorenz curve condition on the interval  $[0, 1]$  to a finite set of independent inequalities and can therefore simplify the computation of relative majorization constraints in practical scenarios.

**Proposition 13.** *Let  $\rho, \rho'$  be any two quantum states with Lorenz curves  $L_{\rho|\sigma}(x), L_{\rho'|\sigma'}(x)$ , where  $\sigma, \sigma'$  are*

*states with positive Wigner distributions. Assume that  $L_{\rho'|\sigma'}(x)$  has  $t$  elbows at locations  $x_1, \dots, x_t$ . Then,  $L_{\rho|\sigma}(x) \geq L_{\rho'|\sigma'}(x)$  for all  $x \in [0, 1]$  iff  $L_{\rho|\sigma}(x_i) \geq L_{\rho'|\sigma'}(x_i)$  for all  $i = 1, \dots, t$ .*

*Proof.* If  $L_{\rho|\sigma}(x) \geq L_{\rho'|\sigma'}(x)$  for all  $x \in [0, 1]$ , then the condition on the elbows follows trivially.

Conversely, assume  $L_{\rho|\sigma}(x_i) \geq L_{\rho'|\sigma'}(x_i)$  for all  $i = 1, \dots, t$ . Suppose on the contrary that  $L_{\rho|\sigma}(x)$  dips below  $L_{\rho'|\sigma'}(x)$  at some point  $x = y$  where  $x_i < y < x_{i+1}$  for some  $i$ . Then this implies

$$\begin{aligned} L_{\rho|\sigma}(x_i) &\geq L_{\rho'|\sigma'}(x_i) \\ L_{\rho|\sigma}(y) &< L_{\rho'|\sigma'}(y) \\ L_{\rho|\sigma}(x_{i+1}) &\geq L_{\rho'|\sigma'}(x_{i+1}). \end{aligned}$$

However, since  $L_{\rho'|\sigma'}(x)$  is linear between  $x_i$  and  $x_{i+1}$ , the above conditions imply that  $L_{\rho|\sigma}(x)$  is not concave, which contradicts Proposition 7. Therefore,  $L_{\rho|\sigma}(x) \geq L_{\rho'|\sigma'}(x)$  for all  $x$  in  $[0, 1]$ .  $\square$

### SUPPLEMENTARY NOTE 3: GENERAL ASPECTS OF LORENZ CURVES FOR NOISY STRANGE STATES

Here we provide background detail as to how the Lorenz curve behaves for a noisy Strange state in the large  $n$  limit.

#### Binomial distributions and error bounds

Consider an experiment consisting of  $n$  trials of throwing a coin with probability  $p$  of landing on heads and  $1-p$  of landing on tails. We denote the sums over an even number  $m$  of successful trials by  $\Phi_+(m; n, p)$  and an odd number  $m$  of successful trials by  $\Phi_-(m; n, p)$ , and they are given by

$$\begin{aligned} \Phi_+(m; n, p) &:= \sum_{\ell=0}^{m/2} \binom{n}{2\ell} p^{2\ell} (1-p)^{n-2\ell}, \\ &\text{for even integers } m \in [0, n], \end{aligned} \quad (24)$$

$$\begin{aligned} \Phi_-(m; n, p) &:= \sum_{\ell=1}^{(m-1)/2} \binom{n}{2\ell+1} p^{2\ell+1} (1-p)^{n-(2\ell+1)}, \\ &\text{for odd integers } m \in [0, n]. \end{aligned} \quad (25)$$

In the next section we will use  $\Phi_+$  and  $\Phi_-$  to express the elbow coordinates of Lorenz curves for unital protocols.

We have the Shannon entropy of a  $p$ -coin and the Shannon relative entropy for a  $p$ -coin and a  $q$ -coin given by the expressions

$$S(p) := -p \log p - (1-p) \log(1-p), \quad (26)$$

$$S(p||q) := p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}. \quad (27)$$

They are symmetric about  $p = 1/2$  and  $q = 1/2$ , namely  $S(p) = S(1-p)$  and  $S(p||q) = S(1-p||1-q)$ . We also have the entropic bounds

**Proposition 14.** *For all  $\ell \in [1, n-1]$ ,*

$$\left[8\ell \left(1 - \frac{\ell}{n}\right)\right]^{-\frac{1}{2}} 2^{nS(\frac{\ell}{n})} \leq \binom{n}{\ell} \leq \quad (28)$$

$$\left[2\pi\ell \left(1 - \frac{\ell}{n}\right)\right]^{-\frac{1}{2}} 2^{nS(\frac{\ell}{n})}. \quad (29)$$

The proof provided in [13] proceeds with direct calculation for the edge cases  $\ell = 1, 2, n-1, n-2$  and use of Stirling's approximation for the remaining cases. The bounds in Proposition 14 can be directly inserted into the expressions of functions  $\Phi_+$  and  $\Phi_-$  to derive strict upper and lower bounds.

It is also possible to derive more manageable bounds on  $\Phi_+, \Phi_-$ . To this end, we rewrite the functions as

$$\Phi_{\pm}(m; n, a) = \frac{1}{2}(\Phi(m; n, a) \pm (1+a)^{-n} S(m; n, a)), \quad (30)$$

where  $a = p/(1-p)$ , and  $\Phi$  is the standard cumulative function for the binomial distribution

$$\Phi(m; n, a) = (1+a)^{-n} \sum_{k=0}^m \binom{n}{k} a^k, \quad (31)$$

and we define the remainder term

$$S(m; n, a) := \sum_{k=0}^m \binom{n}{k} (-a)^k. \quad (32)$$

The cumulative function  $\Phi$  obeys the following bounds, which are proven in [13].

**Proposition 15.** *Given fixed  $n > 0$  and  $p$ ,  $\Phi$  satisfies the following bounds:*

1.  $\Phi(m; n, p) \geq \left[8m \left(1 - \frac{m}{n}\right)\right]^{-\frac{1}{2}} 2^{-nS(\frac{m}{n}||p)},$   
 $m \in [1, n-1]$
2.  $\Phi(m; n, p) \geq 1 - 2^{-nS(\frac{m+1}{n}||p)}, m \in [np+1, n-2]$
3.  $\Phi(m; n, p) \leq 1 - \left[8(m+1) \left(1 - \frac{m+1}{n}\right)\right]^{-\frac{1}{2}} \times$   
 $2^{-nS(\frac{m+1}{n}||p)}, m \in [0, n-2]$
4.  $\Phi(m; n, p) \leq 2^{-nS(\frac{m}{n}||p)}, m \in [0, np]$

We can now put estimates on  $S(m; n, a)$ , and hence the functions  $\Phi_{\pm}$ , for different parameter regimes. We can consider the function  $f(x) = (1+x)^n$  and note that  $S(m; n, a)$  is the  $m$ 'th partial sum of this expansion at the point  $x = -a$ . The truncated Maclaurin series is given by

$$f(x) = f(0) + xf'(0) + \dots + \frac{x^m}{m!} f^{(m)}(0) + R_m(x) \quad (33)$$

with a remainder term

$$R_m(x) = \int_0^x dt f^{(m+1)}(t) \frac{(x-t)^m}{m!} \quad (34)$$

$$= \frac{x^{m+1}}{(m+1)!} f^{(m+1)}(x_*), \quad (35)$$

where in the second expression,  $x_*$  is a point that lies between 0 and  $x$  that comes from the Mean Value Theorem.

Applying this to the function  $f(x) = (1+x)^n$  gives

$$(1+x)^n = \sum_{k=0}^m \binom{n}{k} x^k + R_m. \quad (36)$$

Therefore, if we evaluate at  $x = -a$  we obtain

$$S(m; n, a) = (1-a)^n - R_m(-a), \quad (37)$$

where the remainder term is given by

$$R_m(-a) = \int_0^{-a} dt f^{(m+1)}(t) \frac{(-a-t)^m}{m!} \quad (38)$$

$$= \frac{(-a)^{m+1}}{(m+1)!} f^{(m+1)}(x_*). \quad (39)$$

We can compute the derivative  $f^{(m+1)}(x)$  explicitly,

$$f^{(m+1)}(x) = (m+1)! \binom{n}{m+1} (1+x)^{n-m-1}. \quad (40)$$

Therefore, we have that

$$R_m(-a) = (-1)^{m+1} (m+1) \binom{n}{m+1} \times \int_{-a}^0 dt (1+t)^{n-m-1} (a+t)^m \quad (41)$$

$$= \binom{n}{m+1} (-a)^{m+1} (1+x_*)^{n-m-1}, \quad (42)$$

where in the latter expression  $x_* \in [-a, 0]$ . The first integral expression can be estimated via the Cauchy-Schwarz or the Hölder inequality. Therefore, one has an explicit form with an unknown (but bounded) parameter  $x_*$ , or the integral form.

A very simple estimate, based on  $x_*$  lying in the interval  $[-a, 0]$  gives

$$(1-a)^{n-m-1} \leq \frac{R_m(-a)}{\binom{n}{m+1} (-a)^{m+1}} \leq 1, \quad (43)$$

which in turn leads to the following bounds on  $\Phi_+(m; n, a)$ :

$$2\Phi_+(m; n, a) \leq \Phi(m; n, a) + (1-a)^n - \frac{(-a)^{m+1}}{m!} \times \binom{n}{m+1} (1-a)^{n-m-1} \text{ and} \quad (44)$$

$$2\Phi_+(m; n, a) \geq \Phi(m; n, a) + (1-a)^n - \frac{(-a)^{m+1}}{m!} \times \binom{n}{m+1}. \quad (45)$$

Better bounds can be obtained with a finer analysis, and it would be of interest to obtain tighter asymptotic behaviour on  $\Phi_{\pm}$  that could in turn be used to improve magic distillation bounds.

### Lorenz curve coordinates for unital protocols

The Wigner distribution of the  $n$ -copy qutrit maximally mixed state  $(\mathbb{1}/3)^{\otimes n}$  is the uniform probability distribution over the phase space, consisting of  $9^n$  components equal to  $9^{-n}$ . The Wigner distribution of the 1-copy  $\epsilon$ -noisy Strange state  $\rho_S(\epsilon)$  for unital protocols consists of some permutation of a single negative component

$$-v(\epsilon) := -\left(\frac{1}{3} - \frac{4}{9}\epsilon\right), \quad (46)$$

and 8 positive components each with value

$$u(\epsilon) := \frac{1}{6} - \frac{1}{18}\epsilon. \quad (47)$$

For unital protocols we need the condition  $0 \leq \epsilon < 3/4$ , so that the state contains some Wigner negativity ( $-v < 0$ ). It is also clear that  $v \geq u$  in the interval  $0 \leq \epsilon \leq 3/7$ , while  $u > v$  in the interval  $3/7 < \epsilon < 3/4$ .

The Wigner distribution of the  $n$ -copy  $\epsilon$ -noisy Strange state  $\rho_S(\epsilon)^{\otimes n}$  is given by  $W_{\rho_S(\epsilon)^{\otimes n}} = W_{\rho_S(\epsilon)}^{\otimes n}$  and, using Proposition 11 for unital protocols, we can associate it with a set  $\{(w_i, m_i)\}_{i=0,\dots,n}$  of component-multiplicity pairs, the elements of which are presented in Supplementary Table I for all different combinations of physical parameters  $n, \epsilon$ . We split the values of index  $i$  into two intervals, defined precisely as

$$\text{LHS: } 0 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor \text{ and} \quad (48)$$

$$\text{RHS: } \left\lfloor \frac{n}{2} \right\rfloor + 1 \leq i \leq n. \quad (49)$$

The labels LHS, RHS denote that the components contribute to the left hand side or the right hand side of the Lorenz curve maximum. The maximum is included in the LHS interval.

For unital protocols, the rescaled distribution is simply proportional to the Strange state distribution,

$$W_{\rho_S(\epsilon)^{\otimes n} | (\mathbb{1}/3)^{\otimes n}} = 9^n W_{\rho_S(\epsilon)^{\otimes n}}, \quad (50)$$

therefore expressions for the Lorenz curve coordinates follow from summing up the Wigner components in decreasing order. As a result, every Strange state Lorenz curve for a unital protocol contains  $n$  elbows. In the case of accidental degeneracies among Wigner components, some of these elbows may be straightened out. Including the boundary points  $(x_{-1}, L_{-1}) := (0, 0)$  and  $(x_n, L_n) := (1, 1)$ , we label these elbows as

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

| Case                    |          | $m_i$ | $w_i$                        |
|-------------------------|----------|-------|------------------------------|
| $0 \leq \epsilon < 3/4$ | $n$ even | LHS   | $8^{2i} \binom{n}{2i}$       |
|                         | $n$ even | RHS   | $8^{n-2i} \binom{n}{2i}$     |
|                         | $n$ odd  | LHS   | $8^{2i+1} \binom{n}{2i+1}$   |
|                         | $n$ odd  | RHS   | $8^{n-2i-1} \binom{n}{2i+1}$ |
| $3/7 < \epsilon < 3/4$  | $n$ even | LHS   | $8^{n-2i} \binom{n}{2i}$     |
|                         | $n$ even | RHS   | $8^{2i} \binom{n}{2i}$       |
|                         | $n$ odd  | LHS   | $8^{n-2i} \binom{n}{2i}$     |
|                         | $n$ odd  | RHS   | $8^{2i} \binom{n}{2i}$       |

Supplementary Table I. Wigner components  $w_i(n, \epsilon)$  of  $\rho_S(\epsilon)^{\otimes n}$  along with their multiplicities  $m_i(n, \epsilon)$ , with  $0 \leq i \leq n$ . The expressions change depending on the error  $\epsilon$ , the parity of the number of copies  $n$  and whether the index  $i$  is lower or higher than the index of the Lorenz curve maximum (LHS or RHS). The term  $2i$  is considered mod  $(n+1)$ , so that  $2i \bmod (n+1) = 2i - n - 1$  if  $i > \lfloor n/2 \rfloor$ .

where the coordinates are given by

$$(x_i, L_i) = \left( \frac{1}{9^n} \sum_{\ell=0}^i m_{\ell}, \sum_{\ell=0}^i m_{\ell} w_{\ell} \right). \quad (51)$$

The Lorenz curve maximum is the  $\lfloor n/2 \rfloor$ -th elbow and its coordinates are calculated by collecting all the positive Wigner components,

$$x_{\lfloor n/2 \rfloor} = \frac{1}{2} \left( 1 + \left( \frac{7}{9} \right)^n \right), \quad (52)$$

$$L_{\lfloor n/2 \rfloor} = \frac{1}{2} \left( 1 + \left( \frac{15-8\epsilon}{9} \right)^n \right). \quad (53)$$

In Supplementary Table II, we present the elbow coordinates of the  $n$ -copy,  $\epsilon$ -noisy Strange state Lorenz curve for a unital protocol for any combination of physical parameters  $n, \epsilon$ .

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

$$x_{ij} = \left( 1 - \frac{j}{m_i} \right) x_{i-1} + \frac{j}{m_i} x_i, \quad (54)$$

$$L_{ij} = \left( 1 - \frac{j}{m_i} \right) L_{i-1} + \frac{j}{m_i} L_i \quad (55)$$

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

One can consider the state

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

where tensoring with the maximally mixed state keeps the Lorenz curve unchanged, but increases the resolution

|                                           |          |     |                              |                                                                                           |
|-------------------------------------------|----------|-----|------------------------------|-------------------------------------------------------------------------------------------|
| $0 \leq \epsilon < \frac{3}{4}$           | $n$ even | LHS | $\Phi_+(2i; n, \frac{8}{9})$ | $(\frac{5}{3} - \frac{8}{9}\epsilon)^n \Phi_+(2i; n, \frac{12-4\epsilon}{15-8\epsilon})$  |
|                                           | $n$ odd  | RHS | $\Phi_-(2i; n, \frac{1}{9})$ | $-(\frac{5}{3} - \frac{8}{9}\epsilon)^n \Phi_-(2i; n, \frac{3-4\epsilon}{15-8\epsilon})$  |
| $\frac{3}{4} \leq \epsilon < \frac{1}{2}$ | $n$ even | LHS | $\Phi_-(2i; n, \frac{8}{9})$ | $(\frac{5}{3} - \frac{8}{9}\epsilon)^n \Phi_-(2i; n, \frac{12-4\epsilon}{15-8\epsilon})$  |
|                                           | $n$ odd  | RHS | $\Phi_-(2i; n, \frac{1}{9})$ | $-(\frac{5}{3} - \frac{8}{9}\epsilon)^n \Phi_-(2i; n, \frac{3-4\epsilon}{15-8\epsilon})$  |
| $\frac{1}{2} \leq \epsilon < \frac{3}{4}$ | $n$ even | LHS | $\Phi_+(2i; n, \frac{1}{9})$ | $(\frac{5}{3} - \frac{8}{9}\epsilon)^n \Phi_+(2i; n, \frac{3-4\epsilon}{15-8\epsilon})$   |
|                                           | $n$ odd  | RHS | $\Phi_+(2i; n, \frac{8}{9})$ | $-(\frac{5}{3} - \frac{8}{9}\epsilon)^n \Phi_+(2i; n, \frac{12-4\epsilon}{15-8\epsilon})$ |

Supplementary Table II. Strange state Lorenz curve elbow coordinates for a unital protocol, with  $0 \leq i \leq n$ . For the LHS cases, the  $i$  in the first column labels the elbow coordinate  $x_i$  and the second column gives  $L_i = L(x_i)$ . For RHS cases, the  $i$  in the first column labels  $x_i - x_{\lfloor n/2 \rfloor}$  and the second column gives  $L_i - L_{\lfloor n/2 \rfloor}$ . The coordinate expressions depend on the error  $\epsilon$ , the parity of the number of copies  $n$  and the location of the elbow relative to the maximum (LHS or RHS). The term  $2i$  is considered mod  $(n+1)$ , so that  $2i \bmod (n+1) = 2i - n - 1$  if  $i > \lfloor n/2 \rfloor$ . The point  $(x_{-1}, L_{-1}) := (0, 0)$  is not included in the table.

of (the uniformly distributed) points. The new point coordinates are given by:

$$x_{ijk} = (1 - p_{ijk}) x_{i-1} + p_{ijk} x_i \quad (56)$$

$$L_{ijk} = (1 - p_{ijk}) L_{i-1} + p_{ijk} L_i, \quad (57)$$

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

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

We can also combine the indices, by introducing a single index

$$I(i, j, k) := k + \left[ (j-1) + \sum_{\ell=0}^{i-1} m_\ell(n', \epsilon') \right] 9^{n-n'}, \quad (58)$$

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

$$I(i, m_i(n', \epsilon'), 9^{n-n'}) = \sum_{\ell=0}^i m_\ell(n', \epsilon'), \quad i = 0, \dots, n'. \quad (59)$$

The indexing  $I$  is bijective, i.e.

$$(i, j, k) = (i', j', k') \text{ iff } I(i, j, k) = I(i', j', k'). \quad (60)$$

#### SUPPLEMENTARY NOTE 4: TEMPERATURE-DEPENDENT DISTILLATION BOUNDS FROM RELATIVE MAJORIZATION

Here we prove our main theorem, which bounds magic distillation rates in terms of free energies.

**Theorem 7.** Consider a magic distillation protocol on qutrits that transforms  $n$  copies of an  $\epsilon$ -noisy Strange state into  $m$  copies of an  $\epsilon'$ -noisy Strange state, with depolarising errors  $\epsilon' \leq \epsilon \leq 3/7$ . We also allow pre/post-processing by local Clifford unitaries.

Let  $T = (k\beta)^{-1}$  be any finite temperature for the physical system and let  $H = \sum_{k \in \mathbb{Z}_3} E_k |E_k\rangle \langle E_k|$  be the Hamiltonian of each qutrit subsystem in its eigen-decomposition. Assume that in the thermodynamic limit ( $n, m \gg 1$ ), the protocol applied to the equilibrium state  $\tau^{\otimes n} = (e^{-\beta H} / \mathcal{Z})^{\otimes n}$  maps  $\tau^{\otimes n} \rightarrow \tau'^{\otimes m}$ , where we write  $\tau' = e^{-\beta H'} / \mathcal{Z}'$  for some Hermitian  $H'$ .

Then the asymptotic magic distillation rate  $R = m/n$  is bounded as

$$R \leq \frac{\log(1 - \frac{4}{3}\epsilon) + \beta(\phi - F)}{\log(1 - \frac{4}{3}\epsilon') + \beta(\phi' - F')}, \quad (61)$$

where  $F$  is the free energy of  $\tau$ , and

$$\phi = -\beta^{-1} \log \zeta \quad (62)$$

with  $\zeta$  given by the expressions

$$\zeta = \sum_{k \in \mathbb{Z}_3} \alpha_k e^{-\beta E_k}, \quad (63)$$

$$\alpha_k = \langle E_k | A_{\mathbf{z}_*} | E_k \rangle, \quad (64)$$

and  $W_\tau(\mathbf{z})$  attaining a minimum at  $\mathbf{z} = \mathbf{z}_*$ . The primed variables are defined similarly for the output system.

*Proof.* For the sake of clarity, we write  $\rho_n := \rho_S(\epsilon)^{\otimes n}$ ,  $\rho'_m := \rho_S(\epsilon')^{\otimes m}$ ,  $\tau_n := \tau^{\otimes n}$  and  $\tau'_m := \tau'^{\otimes m}$ . We also assume, without loss of generality on the asymptotic distillation rate, that  $n$  and  $m$  are both even.

To establish the distillation bound we consider the distillation protocol that gives  $\mathcal{E}(\rho_n) = \rho'_m$  for the magic states. We then consider that the protocol transforms the reference equilibrium state as  $\mathcal{E}(\tau_n) = \tau'_m$ . Since we have a finite temperature we have that  $\tau$  and  $\tau'$  are full rank stabilizer states and have a strictly positive Wigner distribution. The input and output magic states generally have quasi-probability Wigner distributions. For any such protocol we therefore have that

$$(W_{\rho_n}(\mathbf{z}), W_{\tau_n}(\mathbf{z})) \succ (W_{\rho'_m}(\mathbf{z}), W_{\tau'_m}(\mathbf{z})), \quad (65)$$

or, equivalently, in terms of the relevant Lorenz curves,

$$L_{\rho_n|\tau_n}(x) \geq L_{\rho'_m|\tau'_m}(x) \text{ for all } x \in [0, 1]. \quad (66)$$

We now consider the rescaled Wigner distribution  $W_{\rho|\tau}(\mathbf{z}) := W_\rho(\mathbf{z})/W_\tau(\mathbf{z})$ , which is well-defined since  $\tau$  is full-rank. Due to the multiplicative property of the Wigner distribution, the rescaled distribution is also multiplicative in the sense that

$$W_{\rho \otimes \rho' | \tau \otimes \tau'}(\mathbf{z}_1 \oplus \mathbf{z}_2) = W_{\rho|\tau}(\mathbf{z}_1) W_{\rho'|\tau'}(\mathbf{z}_2), \quad (67)$$

for any states  $\rho, \rho'$  and any full-rank stabilizer states  $\tau, \tau'$ . Therefore, we have that

$$W_{\rho_n|\tau_n}(\mathbf{z}) = \prod_{i=1}^n W_{\rho|\tau}(\mathbf{z}_i) \quad (68)$$

$$W_{\rho_n}(\mathbf{z}) = \prod_{i=1}^n W_{\rho}(\mathbf{z}_i) \quad (69)$$

where  $\mathbf{z} = \oplus_{i=1}^n \mathbf{z}_i \in \mathbb{Z}_3^{2n}$  is the phase space point for the full system in terms of those of the individual subsystems.

The points defining the Lorenz curve  $L_{\rho_n|\tau_n}(x)$  are obtained from sorting the components of  $W_{\rho_n|\tau_n}(\mathbf{z})$  in non-increasing order and then computing the partial sums of  $W_{\rho_n|\tau_n}(\pi(\mathbf{z}))$  where  $\pi$  is the permutation that realizes the sorting. However, similar to the unital protocol analysis, we use the slope constraint (Proposition 12) obtained by considering the line segments connecting the origin to the first elbow of both Lorenz curves.

The Wigner distribution of a single noisy Strange state consists of 1 negative component  $W_{\rho}(\mathbf{0}) = -v(\epsilon)$  and 8 positive components  $W_{\rho}(\mathbf{z}) = u(\epsilon)$  for  $\mathbf{z} \neq \mathbf{0}$ . The Wigner distribution of the full-rank, stabilizer equilibrium state  $\tau$  is  $W_{\tau}(\mathbf{z}) > 0$  for all  $\mathbf{z} \in \mathbb{Z}_3^2$ .

Assume that the smallest component of the distribution  $W_{\tau}(\mathbf{z})$  is at  $\mathbf{z} = \mathbf{z}_*$ . Since the magic content of the state is unchanged under a Clifford unitary  $C$ , we can instead consider the state

$$\rho_S(\epsilon) \rightarrow C\rho_S(\epsilon)C^\dagger = D_{\mathbf{z}_*}\rho_S(\epsilon)D_{\mathbf{z}_*}^\dagger, \quad (70)$$

which has its single negative Wigner component  $-v(\epsilon)$  at the point  $\mathbf{z}_*$ .

The components of the rescaled distribution  $W_{\rho_n|\tau_n}$  for this transformed state are given by

$$\left(\frac{-v}{W_{\tau}(\mathbf{z}_*)}\right)^{i_{\mathbf{z}_*}} \prod_{\mathbf{z} \neq \mathbf{z}_*} \left(\frac{u}{W_{\tau}(\mathbf{z})}\right)^{i_{\mathbf{z}}}, \quad (71)$$

where the integer indices obey the following conditions:

$$\begin{aligned} 0 \leq i_{\mathbf{z}} \leq n \text{ for all } \mathbf{z} \in \mathbb{Z}_3^2, \\ \sum_{\mathbf{z} \in \mathbb{Z}_3^2} i_{\mathbf{z}} = n. \end{aligned} \quad (72)$$

We now compute the largest rescaled component. Firstly, note that  $n$  is even, so we require that  $i_{\mathbf{z}_*} \in \{0, 2, \dots, n\}$  for the component to be positive. Then, we have that  $v \geq u$  because  $\epsilon \leq 3/7$ , and we have already ensured that  $W_{\tau}(\mathbf{z}_*) \leq W_{\tau}(\mathbf{z})$  for all  $\mathbf{z} \in \mathbb{Z}_3^2$ . Therefore, the largest rescaled component occurs when  $i_{\mathbf{z}_*} = n$  and  $i_{\mathbf{z}} = 0$  for  $\mathbf{z} \neq \mathbf{z}_*$  and is equal to  $(v/W_{\tau}(\mathbf{z}_*))^n$ . The coordinates of the first Lorenz curve point after the origin are given by

$$(x_0, L_0) = ((W_{\tau}(\mathbf{z}_*))^n, v^n) \quad (73)$$

Given a Hamiltonian decomposition,

$$H = \sum_{k \in \mathbb{Z}_3} E_k |E_k\rangle\langle E_k|, \quad (74)$$

we now express the coordinates of the first point in terms of free energy quantities. We expand as follows

$$\begin{aligned} W_{\tau}(\mathbf{z}_*) &= \frac{1}{3\mathcal{Z}} \text{tr} [A_{\mathbf{z}_*} e^{-\beta H}] \\ &= \frac{e^{\beta F}}{3} \text{tr} \left[ A_{\mathbf{z}_*} \sum_{k \in \mathbb{Z}_3} e^{-\beta E_k} |E_k\rangle\langle E_k| \right] \\ &= \frac{e^{\beta F}}{3} \sum_{r \in \mathbb{Z}_3} \alpha_r e^{-\beta E_r} = \frac{e^{\beta F}}{3} \zeta \\ &= \frac{e^{\beta(F-\phi)}}{3}, \end{aligned} \quad (75)$$

where we define

$$\alpha_k := \langle E_k | A_{\mathbf{z}_*} | E_k \rangle \quad (76)$$

$$\zeta := \sum_{k \in \mathbb{Z}_3} \alpha_k e^{-\beta E_k}, \quad (77)$$

and a ‘magic free energy’ term  $\phi := -\beta^{-1} \log \zeta$ . Therefore, the coordinates can be expressed as

$$(x_0, L_0) = \left( \frac{e^{n\beta(F-\phi)}}{3^n}, v(\epsilon)^n \right). \quad (78)$$

The output states obey the same conditions, so we also have the first point coordinates of the output Lorenz curve given by

$$(x'_0, L'_0) = \left( \frac{e^{m\beta(F'-\phi')}}{3^m}, v(\epsilon')^m \right). \quad (79)$$

If the largest rescaled component of a state is distinct with no multiplicities, then these coordinates correspond to the first elbow of the corresponding Lorenz curve, whereas if it appears multiple times, then the coordinates derived correspond to a point on the interior of the line segment connecting the origin to the first elbow. In both cases, the distillation bound remains the same, as is clear by its derivation in Proposition 12, given by  $L_0/x_0 \geq L'_0/x'_0$ , leading to

$$\left( 3v(\epsilon)e^{-\beta(F-\phi)} \right)^n = \frac{L_0}{x_0} \geq \frac{L'_0}{x'_0} = \left( 3v(\epsilon')e^{-\beta(F'-\phi')} \right)^m. \quad (80)$$

We note that  $3v(\epsilon')e^{-\beta(F'-\phi')} > 1$  always, as  $\rho_S(\epsilon')$  is not a free state, so the initial slope of its Lorenz curve exceeds 1. Taking the natural logarithm on both sides and rearranging gives the bound

$$\frac{m}{n} \leq \frac{\log(1 - \frac{4}{3}\epsilon) + \beta(\phi - F)}{\log(1 - \frac{4}{3}\epsilon') + \beta(\phi' - F')}, \quad (81)$$

which completes the proof.  $\square$

The following result shows that if the energy eigenbasis is a stabilizer basis then the above bounds simplify further.

**Proposition 16.** *Let  $H$  be a Hamiltonian  $H$  with spectrum  $\{E_k\}_{k \in \mathbb{Z}_d}$ . If  $H$  has a stabilizer eigenbasis, then  $\phi = E_{k_*}$  for some  $k_* \in \mathbb{Z}_d$ .*

*Proof.* We first write  $\alpha_k$  as

$$\alpha_k = \langle E_k | A_{\mathbf{z}_*} | E_k \rangle = dW_{|E_k\rangle\langle E_k|}(\mathbf{z}_*) \quad (82)$$

Since each  $|E_k\rangle\langle E_k|$  is a stabilizer state, their Wigner distributions have orthogonal supports in the phase space, so only one of them, say  $k = k_*$  contains point  $\mathbf{z}_*$  in its support. Additionally, its distribution is uniform on its support, so  $W_{|E_{k_*}\rangle\langle E_{k_*}|}(\mathbf{z}_*) = 1/d$ .

Therefore, only the coefficient  $\alpha_{k_*}$  is non-zero and

$$\begin{aligned} \phi &= -\beta^{-1} \log \left( \sum_k \alpha_k e^{-\beta E_k} \right) \\ &= E_{k_*} - kT \log(dW_{|E_{k_*}\rangle\langle E_{k_*}|}(\mathbf{z}_*)) = E_{k_*}. \end{aligned} \quad (83)$$

□

### Dropping pre/post-processing of the magic states

If we drop the freedom to pre/post-process the magic states via Clifford unitaries, then the final expression for the bounds change and become a little more complex, since the largest component in the rescaled distribution depends on a number of factors. We illustrate this by the following result.

**Proposition 17.** *Consider a magic distillation protocol on qutrits that transforms*

$$\rho_S(\epsilon)^{\otimes n} \longrightarrow \mathcal{E}(\rho_S(\epsilon)^{\otimes n}) = \rho_S(\epsilon')^{\otimes m}$$

with  $n, m \gg 1$ .

Let each qutrit have a Hamiltonian  $H$  with energies  $E_0, E_1, E_2$  and stabilizer eigenstates, and define  $E_{\max} = \max\{E_0, E_1, E_2\}$  and  $E_s$  as the eigenvalue of the Hamiltonian eigenstate whose Wigner distribution overlaps the negative component of  $\rho_S(\epsilon)$  in the phase space. Let  $T = (k\beta)^{-1}$  be any finite temperature and assume that in the thermodynamic limit ( $n, m \gg 1$ ), the protocol applied to the equilibrium state  $\tau^{\otimes n} = (e^{-\beta H}/\mathcal{Z})^{\otimes n}$  maps  $\tau^{\otimes n} \longrightarrow \tau'^{\otimes m}$ , where we express state  $\tau'$  as  $\tau' = e^{-\beta H'}/\mathcal{Z}'$  for some Hermitian  $H'$ .

Define  $\beta_* = (kT_*)^{-1}$  through the relation

$$E_{\max} - E_s =: kT_* \ln 2, \quad (84)$$

and define a threshold error,

$$\epsilon_*(\beta) := \begin{cases} 3 - \frac{9}{4 - 2^{\beta/\beta_* - 1}}, & \text{for } \beta \leq \beta_* \\ 0, & \text{for } \beta > \beta_*. \end{cases} \quad (85)$$

Then, the distillation rate  $R = m/n$  of the magic protocol is bounded as:

$$R \leq \begin{cases} \frac{\ln(1 - \frac{4}{3}\epsilon) + \beta(E_s - F)}{\ln(1 - \frac{4}{3}\epsilon') + \beta(E'_s - F')}, & \epsilon \leq \epsilon_*, \epsilon' \leq \epsilon'_*, \\ \frac{\ln(1 - \frac{4}{3}\epsilon) + \beta(E_s - F)}{\ln(\frac{1}{2} - \frac{1}{6}\epsilon') + \beta(E'_{\max} - F')}, & \epsilon \leq \epsilon_*, \epsilon' > \epsilon'_*, \\ \frac{\ln(\frac{1}{2} - \frac{1}{6}\epsilon) + \beta(E_{\max} - F)}{\ln(1 - \frac{4}{3}\epsilon') + \beta(E'_s - F')}, & \epsilon > \epsilon_*, \epsilon' \leq \epsilon'_*, \\ \frac{\ln(\frac{1}{2} - \frac{1}{6}\epsilon) + \beta(E_{\max} - F)}{\ln(\frac{1}{2} - \frac{1}{6}\epsilon') + \beta(E'_{\max} - F')}, & \epsilon > \epsilon_*, \epsilon' > \epsilon'_*, \end{cases} \quad (86)$$

where  $F$  is the free energy of state  $\tau$  and all primed quantities are similarly defined for the output system.

*Proof.* The proof proceeds in the same manner as in Theorem 7 up to the point where we need to evaluate  $W_{\rho_n|\tau_n}(\mathbf{z})$  and  $W_{\rho_n}(\mathbf{z})$  explicitly, beyond which we do not have the freedom of additional Clifford processing.

The equilibrium state at inverse temperature  $\beta$  on a single qutrit is given by  $\tau = e^{-\beta H}/\mathcal{Z}$ . Moreover, we have that  $\tau$  is a full-rank stabilizer state, where  $\beta \geq 0$  and  $H = E_0|\varphi_0\rangle\langle\varphi_0| + E_1|\varphi_1\rangle\langle\varphi_1| + E_2|\varphi_2\rangle\langle\varphi_2|$  is an eigendecomposition of  $H$ . The state  $\tau$  can now be written as

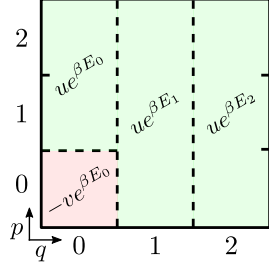
$$\tau = \frac{e^{-\beta E_0}}{\mathcal{Z}}|\varphi_0\rangle\langle\varphi_0| + \frac{e^{-\beta E_1}}{\mathcal{Z}}|\varphi_1\rangle\langle\varphi_1| + \frac{e^{-\beta E_2}}{\mathcal{Z}}|\varphi_2\rangle\langle\varphi_2|, \quad (87)$$

where the eigenstates  $\{|\varphi_k\rangle\langle\varphi_k|\}$  are pure, orthonormal stabilizer states, which can be represented in terms of generalized Paulis. We are free to redefine the computational basis so it coincides with the basis of  $H$ . Abstractly, let  $C$  be the unitary transforming each  $|\varphi_k\rangle\langle\varphi_k|$  to  $|k\rangle\langle k|$ . Since the Clifford group is the normalizer of the Heisenberg-Weyl group,  $C$  is a Clifford unitary. Therefore,  $C$  maps  $\tau$  to another stabilizer state that is diagonal in the computational basis, and we can assume without loss of generality that  $\tau$  is diagonal in  $\{|0\rangle, |1\rangle, |2\rangle\}$ . This re-definition of the computational basis means that the coordinates on the discrete phase space are also permuted, so the negative Wigner component  $-v(\epsilon)$  of the Strange state is in a new position. We denote by  $E_s$  the eigenvalue of  $H$  where the associated eigenvector has Wigner distribution overlapping the negative component of the state  $C\rho_S(\epsilon)C^\dagger$  in the original phase space, hence also of the state Strange state  $\rho_S(\epsilon)$  in the re-defined phase space. The eigenvalue  $E_s$  is unique, since the eigenstates form an orthonormal basis.

The Wigner distribution of state  $\tau$  is then given by

$$\begin{aligned} W_\tau(\mathbf{z}) &= \sum_{k=0}^2 \frac{e^{-\beta E_k}}{\mathcal{Z}} W_{|k\rangle\langle k|}(\mathbf{z}) \\ &= \sum_{k=0}^2 \frac{e^{-\beta E_k}}{\mathcal{Z}} \delta_{\mathbf{z}, \mathbf{k}} = \frac{e^{-\beta E_q}}{3\mathcal{Z}}, \end{aligned} \quad (88)$$

where  $x$  labels one of the three vertical lines in the phase space. The rescaled Wigner distribution  $W_{\rho|\tau}(\mathbf{z})$  can now be computed. It has 9 components, but these come with multiplicities. In total, there are four distinct values on the phase space, as illustrated in Supplementary Figure 1.



Supplementary Figure 1. **Qutrit phase space regions for  $W_{\rho|\tau}(\mathbf{z})$ .** Here, the negative component of the magic state overlaps the Wigner distribution of  $|0\rangle$ . The rescaled distribution attains a single value in each of the four regions, proportional to the value depicted in the region, see Eq. (92).

We define vectors  $\mathbf{w}(\rho)$ ,  $\mathbf{w}(\tau)$ ,  $\mathbf{w}(\rho|\tau)$  based on the values occurring in  $W_\rho$ ,  $W_\tau$ ,  $W_{\rho|\tau}$  respectively and  $\mathbf{m}$  as the vector of associated multiplicities of each value in  $W_\rho(\mathbf{z})$ . Specifically, the component values and multiplicities of the relevant distributions in the four distinct regions are given by

$$\mathbf{m} := (1, 2, 3, 3), \quad (89)$$

$$\mathbf{w}(\rho) := (-v, u, u, u), \quad (90)$$

$$\mathbf{w}(\tau) := \frac{1}{3\mathcal{Z}} (e^{-\beta E_0}, e^{-\beta E_0}, e^{-\beta E_1}, e^{-\beta E_2}), \quad (91)$$

$$\mathbf{w}(\rho|\tau) := 3\mathcal{Z} (-v e^{\beta E_0}, u e^{\beta E_0}, u e^{\beta E_1}, u e^{\beta E_2}). \quad (92)$$

Using this notation, the values and multiplicities of the  $n$ -copy distribution  $\mathbf{w}(\rho_n|\tau_n)$  are computed in Proposition 11. The values are given by

$$[\mathbf{w}(\rho_n|\tau_n)]_{ijk} = (3\mathcal{Z})^n (-v)^{n-\alpha} u^\alpha e^{\beta(n-\alpha)E_s} e^{\beta(iE_0+jE_1+kE_2)}, \quad (93)$$

where the indices  $i, j, k$  are non-negative integers that obey the constraint  $\alpha := i + j + k \leq n$ . The multiplicity of this above value is  $m_{ijk}$  with

$$m_{ijk} = \frac{n!}{i!j!k!(n-\alpha)!} 2^i 3^j 3^k. \quad (94)$$

The associated components of  $\mathbf{w}(\rho_n)$  are given by

$$[\mathbf{w}(\rho_n)]_{ijk} = (-v)^{n-\alpha} u^\alpha, \quad (95)$$

$$[\mathbf{w}(\tau)]_{ijk} = (3\mathcal{Z})^{-n} e^{-\beta(n-\alpha)E_s} e^{-\beta(iE_0+jE_1+kE_2)}. \quad (96)$$

In order to construct the  $n$ -copy Lorenz curve  $L_{\rho_n|\tau_n}(x)$  we need to order the components of the distribution,  $w(\rho|\tau)_{ijk}$  in decreasing order, and identify the sequence of indices that give us  $W_{\rho_n}(\pi(\mathbf{z}))$ .

However, in order to obtain a non-trivial bound, it again suffices to use the constraint from the first elbow  $(x_0, L_0)$  of  $L_{\rho_n|\tau_n}(\mathbf{z})$ . We therefore compute the largest component

$$w_{\max} := \max_{i,j,k} [\mathbf{w}(\rho_n|\tau_n)]_{ijk}, \quad (97)$$

and determine the indices at which this occurs. Putting in the values we obtain

$$(3\mathcal{Z})^{-n} w_{\max} = \max_{i,j,k} \left\{ (-v)^{n-\alpha} u^\alpha e^{\beta(n-\alpha)E_s} e^{\beta(iE_0+jE_1+kE_2)} \right\}, \quad (98)$$

where  $0 \leq i, j, k \leq n$  and  $\alpha := i + j + k \leq n$ . Now for  $0 \leq \epsilon \leq 3/7$ , we have  $v \geq u$ . Since we assume that  $n$  is even, we need the sum  $\alpha = i + j + k$  to be even too, so that the expression is positive.

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

$$w_{\max} = (3\mathcal{Z})^n v^n e^{n\beta E_s} \max_{\substack{\alpha=0,2, \\ \dots, n-2,n}} \left\{ \left( \frac{u}{v} e^{\beta(E_{\max}-E_s)} \right)^\alpha \right\}. \quad (99)$$

If the expression  $\frac{u(\epsilon)}{v(\epsilon)} e^{\beta(E_{\max}-E_s)}$  is less than 1 then the maximum occurs at  $\alpha = 0$ , otherwise it occurs at  $\alpha = n$ . For a fixed state  $\tau$ , this transition is determined by the value of the depolarising error parameter  $\epsilon$  of the noisy magic state. The transition occurs at  $\epsilon = \epsilon_\star$  where

$$\frac{u(\epsilon_\star)}{v(\epsilon_\star)} e^{\beta(E_{\max}-E_s)} = \frac{3 - \epsilon_\star}{6 - 8\epsilon_\star} e^{\beta(E_{\max}-E_s)} = 1. \quad (100)$$

If  $E_{\max} = E_s$ , 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  $\epsilon_\star = 3/7$ . However, the condition that  $\epsilon_\star \geq 0$  also implies a constraint on the effective temperature of the stabilizer state. Specifically, there is a threshold temperature value  $\beta_\star$  given by

$$\beta_\star := \frac{1}{E_{\max} - E_s} \ln 2, \quad (101)$$

such that for the regime  $0 \leq \beta \leq \beta_\star$  a threshold error  $\epsilon_\star$  exists, and for  $\beta > \beta_\star$  no such transition exists, so we choose  $\epsilon_\star = 0$ . Therefore, the transition value for the error is given by

$$\epsilon_\star(\beta) := \begin{cases} 3 - \frac{9}{4 - 2^{\beta/\beta_\star - 1}}, & \text{for } \beta \leq \beta_\star \\ 0, & \text{for } \beta > \beta_\star. \end{cases} \quad (102)$$

The quantity  $w(\rho_S|\sigma)_{\max}$  is now given by

$$w_{\max} = \begin{cases} (3\mathcal{Z})^n v^n e^{n\beta E_s}, & \text{if } \epsilon \leq \epsilon_\star, \quad (\text{C1}) \\ (3\mathcal{Z})^n u^n e^{n\beta E_{\max}}, & \text{if } \epsilon > \epsilon_\star. \quad (\text{C2}) \end{cases}$$

Case (C1) corresponds to  $(i, j, k) = (0, 0, 0)$ , so the multiplicity is  $m_{000} = 1$ , while Case (C2) corresponds to

$$(i, j, k) = \begin{cases} (0, n, 0), & \text{if } E_{\max} = E_1, \\ (0, 0, n), & \text{if } E_{\max} = E_2, \end{cases} \quad (103)$$

so the multiplicity in both cases is  $3^n$ .

Using that  $F = -\beta^{-1} \log \mathcal{Z}$ , the first elbow coordinates in the two cases are now given by

$$(x_0, L_0) = \begin{cases} \left( \frac{1}{3^n} e^{-n\beta(E_s - F)}, v^n \right), & \epsilon \leq \epsilon_\star \\ \left( e^{-n\beta(E_{\max} - F)}, (3u)^n \right), & \epsilon > \epsilon_\star \end{cases} \quad (104)$$

Similarly, considering the output magic state with respect to state  $\sigma'$ , the image of equilibrium state  $\sigma$  under the magic protocol, we get output Lorenz curve coordinates,

$$(x'_0, L'_0) = \begin{cases} \left( \frac{1}{3^{n'}} e^{-n'\beta(E'_s - F')}, v(\epsilon')^{n'} \right), & \epsilon' \leq \epsilon'_\star \\ \left( e^{-n'\beta(E'_{\max} - F')}, (3u(\epsilon'))^{n'} \right), & \epsilon' > \epsilon'_\star \end{cases} \quad (105)$$

There are four combinations of coordinates, depending on the error parameters  $\epsilon, \epsilon'$  for the input and output states. In each of these combinations, we simply use the first elbow constraint, as described in Proposition 12, and manipulate the coordinates as in the proof of Theorem 7, leading to the bounds in the statement of this theorem.  $\square$

### SUPPLEMENTARY NOTE 5: SINGLE-SHOT ENTROPIES ON QUASI-DISTRIBUTIONS AND MAGIC DISTILLATION BOUNDS

In this section, we analyse Rényi entropies on quasi-distributions and use them to get magic distillation bounds.

#### Rényi min-divergence from Lorenz curves

We first demonstrate that the initial slope analysis on the Wigner quasi-distributions makes use of the min-relative divergence. We define

$$D_\infty(W_\rho || W_\tau) := \log \max_{\mathbf{z}} \frac{W_\rho(\mathbf{z})}{W_\tau(\mathbf{z})}, \quad (106)$$

which is well-defined since both the nominator and denominator in the above expression are strictly positive when  $\tau$  is in the interior of  $\mathcal{F}$ .

**Theorem 8.** *Let  $\tau$  be in the interior of  $\mathcal{F}$ . Then  $D_\infty(W_\rho || W_\tau)$  is well-defined for all  $\rho$ , and the following hold:*

1.  $D_\infty(W_\rho || W_\tau) \geq 0$  for all quantum states  $\rho$ .

2.  $D_\infty(W_\rho || W_\tau) = 0$  if and only if  $\rho = \tau$ .

3.  $D_\infty(W_{\rho^{\otimes 2n}} || W_{\tau^{\otimes 2n}}) = n D_\infty(W_{\rho^{\otimes 2}} || W_{\tau^{\otimes 2}})$  for any  $n \in \mathbb{N}$ .

4.  $D_\infty(W_\rho || W_\tau) \geq D_\infty(W_{\mathcal{E}(\rho)} || W_{\mathcal{E}(\tau)})$  for any free operation  $\mathcal{E}$  such that  $\mathcal{E}(\tau)$  is in the interior of  $\mathcal{F}$ .

*Proof.* Since  $\tau$  is in the interior of  $\mathcal{F}$ , its Wigner function obeys  $W_\tau(\mathbf{z}) > 0$  for all points  $\mathbf{z}$  in the phase space. In general,  $W_\rho(\mathbf{z})$  is a quasi-distribution, but the given form of  $\alpha$  ensures that  $W_\rho(\mathbf{z})^\alpha \geq 0$ . Therefore  $D_\infty(W_\rho || W_\tau)$  is always well-defined.

1. We have that

$$D_\infty(W_\rho || W_\tau) = \log \max_{\mathbf{z}} \frac{W_\rho(\mathbf{z})}{W_\tau(\mathbf{z})}, \quad (107)$$

so  $2^{D_\infty(W_\rho || W_\tau)}$  equals the slope of the Lorenz curve  $L_{\rho|\tau}(x)$  at  $x = 0$ . Since  $L_{\rho|\tau}(x)$  is a concave function passing through  $(0, 0)$  and  $(1, 1)$  this implies that  $L_{\rho|\tau}(x) \geq x$  for all  $x \in [0, 1]$ , and in particular its slope at  $x = 0$  is always greater than or equal to 1, which implies  $D_\infty(W_\rho || W_\tau) \geq 0$  for all  $\rho$ .

2.  $D_\infty(W_\rho || W_\tau) = 0$  implies that  $\max_{\mathbf{z}} \frac{W_\rho(\mathbf{z})}{W_\tau(\mathbf{z})} = 1$  and the initial slope of  $L_{\rho|\tau}(x)$  is 1. From the concavity of the function and the fact that  $L_{\rho|\tau}(1) = 1$ , this implies that the slope of  $L_{\rho|\tau}(x)$  must equal 1 throughout the interval  $[0, 1]$  and this, together with the definition of the Lorenz curve, implies that  $W_\rho(\mathbf{z})/W_\tau(\mathbf{z}) = 1$  for all  $\mathbf{z}$ . Since the Wigner representation  $\rho \mapsto W_\rho(\mathbf{z})$  is bijective this implies that  $\rho = \tau$  only. The converse holds by inspection.

3. Given a vector  $\mathbf{w} \in \mathbb{R}^N$ , it is generally the case that  $\max_k w_k \neq \max_k |w_k|$ . However, we do have that

$$\max_{k_1, k_2} (w_{k_1} w_{k_2}) = \max_{k_1, k_2} |w_{k_1} w_{k_2}|, \quad (108)$$

and if additionally  $w_k \geq 0$  for all  $k$ , then we also have for  $2n$  copies that

$$\max_{k_1, k_2, \dots, k_{2n}} (w_{k_1} w_{k_2} \cdots w_{k_{2n}}) = \left( \max_k |w_k| \right)^{2n}, \quad (109)$$

and also,

$$\begin{aligned} \max_{k_1, k_2, \dots, k_{2n}} (w_{k_1} w_{k_2} \cdots w_{k_{2n}}) &= \\ \max_{k_1, k_2, \dots, k_{2n}} (|w_{k_1} w_{k_2}| \cdots |w_{k_{2n-1}} w_{k_{2n}}|) &= \\ \left( \max_{k_1, k_2} (w_{k_1} w_{k_2}) \right)^n. \end{aligned} \quad (110)$$

If we now let  $\mathbf{w} = W_{\rho|\tau}(\mathbf{z}) := W_\rho(\mathbf{z})/W_\tau(\mathbf{z})$  we have

$$\max_{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{2n}} \prod_{k=1}^{2n} W_{\rho|\tau}(\mathbf{z}_k) = \left( \max_{\mathbf{z}_1, \mathbf{z}_2} W_{\rho|\tau}(\mathbf{z}_1) W_{\rho|\tau}(\mathbf{z}_2) \right)^n \quad (111)$$

and therefore taking logarithms we have that  $D_\infty(W_{\rho^{\otimes 2n}} || W_{\tau^{\otimes 2n}}) = n D_\infty(W_{\rho^{\otimes 2}} || W_{\tau^{\otimes 2}})$  as required.

4. The free operation  $\mathcal{E}$  in the Wigner representation corresponds to a stochastic map, which sends  $W_\tau(\mathbf{z})$  to  $W_{\mathcal{E}(\tau)}(\mathbf{z})$ , which is another strictly positive probability distribution on the phase space, due to  $\mathcal{E}(\tau)$  being in the interior of  $\mathcal{F}$ . Moreover,  $(W_\rho, W_\tau) \succ (W_{\mathcal{E}(\rho)}, W_{\mathcal{E}(\tau)})$ . As shown in the main text this condition holds if and only if  $L_{\rho|\tau}(x) \geq L_{\mathcal{E}(\rho)|\mathcal{E}(\tau)}(x)$  for all  $x$ . In particular, this implies the slope at the origin of the input curve is never less than the slope at the origin of the output curve, and hence the result follows.  $\square$

### Well-defined Rényi entropies on quasi-distributions

We now make use of known results from majorization theory to establish the Schur-concavity of the Rényi entropy  $H_\alpha$  on Wigner quasi-distributions for a subset of  $\alpha$  values. We consider the set of all quasi-distributions that is a hyper-plane in  $\mathbb{R}^N$  given by

$$\mathcal{Q}_N = \left\{ \mathbf{w} \in \mathbb{R}^N : \sum_{i=1}^N w_i = 1 \right\}. \quad (112)$$

We define the notion of Schur-convex and Schur-concave functions on a subset  $\mathcal{D} \subseteq \mathbb{R}^N$ .

**Definition 18.** A continuous, real-valued function  $f$  defined on  $\mathcal{D} \subseteq \mathbb{R}^N$  is called Schur-convex on the domain  $\mathcal{D}$  if  $\mathbf{p} \prec \mathbf{q}$  on  $\mathcal{D}$  implies that  $f(\mathbf{p}) \leq f(\mathbf{q})$ . A continuous function  $f : \mathcal{D} \rightarrow \mathbb{R}$  is called Schur-concave if  $(-f)$  is Schur-convex.

For any subset  $\mathcal{D} \subseteq \mathbb{R}^N$  we define

$$\mathcal{D}^\downarrow := \{ \mathbf{w} \in \mathcal{D} : w_1 \geq w_2 \geq \dots \geq w_N \}. \quad (113)$$

We now have the following fundamental result in majorization theory.

**Proposition 19** (Schur-Ostrowski theorem [9, 14, 15]). Let  $\mathcal{D} \subseteq \mathbb{R}^N$  be a convex set with non-empty interior, and invariant under permutations of vector components. Let  $f : \mathcal{D} \rightarrow \mathbb{R}$  be a continuously differentiable function on the interior of  $\mathcal{D}$  and continuous on the whole of  $\mathcal{D}$ . Then  $f$  is Schur-convex on  $\mathcal{D}$  if and only if  $f$  is a symmetric function on  $\mathcal{D}$  and

$$\partial_1 f(\mathbf{w}) \geq \partial_2 f(\mathbf{w}) \geq \dots \geq \partial_N f(\mathbf{w}), \quad (114)$$

for all  $\mathbf{w}$  in the interior of  $\mathcal{D}^\downarrow$ .

Since the set  $\mathcal{Q}_N$  of quasi-distributions obey the necessary conditions, we get the following result:

**Proposition 20.** Let  $f$  be a real-valued continuous function defined on  $\mathcal{Q}_N$  that is continuously differentiable on the interior of  $\mathcal{Q}_N$ . Then,  $f$  is Schur-convex on  $\mathcal{Q}_N$  if and only if it is a symmetric function and

$$\partial_1 f(\mathbf{w}) \geq \partial_2 f(\mathbf{w}) \geq \dots \geq \partial_N f(\mathbf{w}), \quad (115)$$

for all quasi-distributions  $\mathbf{w}$  in the interior of  $\mathcal{Q}_N^\downarrow$ .

The classical  $\alpha$ -Rényi entropy on a probability distribution  $\mathbf{p} = (p_i)$  is given by

$$H_\alpha(\mathbf{p}) := \frac{1}{1-\alpha} \log \sum_i p_i^\alpha, \quad (116)$$

where  $\alpha \geq 0$ . This means the function is well-defined on  $\mathcal{Q}_N^+$ , namely the set of quasi-distributions with non-negative components. We wish to extend the domain to the whole of  $\mathcal{Q}_N$ , but this requires that  $w_i^\alpha$  be a real number for any  $w_i \in \mathbb{R}$ .

The exponent function  $x \mapsto x^\alpha$  is well-defined and real-valued if  $\alpha$  is a rational number of the form  $\alpha = r/s$  where  $s$  is an odd integer and  $r \in \mathbb{N}$ . However, we also require that  $\sum_i w_i^\alpha > 0$  in order for the logarithm to return a real value. To ensure this, we restrict to  $\alpha = r/s$  with  $r = 2a$  being an even integer and  $s = 2b-1$  being an odd integer, for some integers  $a, b \in \mathbb{N}$ , which implies that the following are all well-defined, real-valued expressions,

$$0 \leq w_i^\alpha = w_i^{\frac{2a}{2b-1}} = \left( w_i^{\frac{1}{2b-1}} \right)^{2a} = (w_i^{2a})^{\frac{1}{2b-1}} = |w_i|^{\frac{2a}{2b-1}}.$$

We note that previous work [16] has considered the  $\alpha = 2$  entropy of Wigner distributions, and other work exists that uses the Wehrl entropy, based on the Hussimi function of a quantum state [17].

We next show that  $H_\alpha(\mathbf{w})$  for  $\alpha = 2a/(2b-1)$  is a Schur-concave function on  $\mathcal{Q}_N$  provided that  $a \geq b$ .

**Theorem 9.** If  $\alpha = \frac{2a}{2b-1}$  for positive integers  $a, b$  with  $a \geq b$ , then  $H_\alpha(\mathbf{w})$  is well-defined on the set of quasi-distributions  $\mathcal{Q}_N$ , and moreover if  $\mathbf{w} \succ \mathbf{w}'$  for two quasi-distributions  $\mathbf{w}, \mathbf{w}' \in \mathcal{Q}_N$  then  $H_\alpha(\mathbf{w}) \leq H_\alpha(\mathbf{w}')$ .

*Proof.* We have  $\alpha = \frac{2a}{2b-1} \geq \frac{2b}{2b-1} > 1$  and  $\alpha$  is a rational with even numerator and odd denominator, so  $\sum_i w_i^\alpha > 0$  and  $H_\alpha(\mathbf{w})$  is well-defined for all quasi-distributions  $\mathbf{w}$ .

We also have that  $H_\alpha(\mathbf{w})$  is well-defined, continuous, differentiable on the interior of  $\mathcal{Q}_N$  and symmetric in the components of  $\mathbf{w}$ .

Consider the partial derivatives

$$\frac{\partial H_\alpha}{\partial w_i} = \frac{\alpha}{(\alpha-1) \sum_k w_k^\alpha} (-w_i^{\alpha-1}). \quad (117)$$

For  $\alpha = \frac{2a}{2b-1}$ , with  $a \geq b$ , we have that

$$\frac{\alpha}{(\alpha-1) \sum_k w_k^\alpha} > 0. \quad (118)$$

The first derivative of the function  $g(w) := -w^{\alpha-1}$  is given by

$$g'(w) = -(\alpha-1)w^{\alpha-2} = -(\alpha-1)w^{\frac{2a-4b+2}{2b-1}} \leq 0, \quad (119)$$

because  $\alpha > 1$  and

$$w^{\frac{2a-4b+2}{2b-1}} = \left( w^{\frac{a-2b+1}{2b-1}} \right)^2 \geq 0, \quad (120)$$

so  $g(w)$  is non-increasing in  $w$ .

Therefore, whenever  $w_i \geq w_j$ , we have that

$$-w_i^{\alpha-1} \leq -w_j^{\alpha-1}, \quad (121)$$

which implies that

$$\frac{\partial H_\alpha}{\partial w_i}(\mathbf{w}) \leq \frac{\partial H_\alpha}{\partial w_j}(\mathbf{w}), \quad (122)$$

for any  $\mathbf{w}$  in the interior of  $\mathcal{Q}_N^\downarrow$ . Therefore,  $H_\alpha$  is Schur-concave on  $\mathcal{Q}_N$  and

$$H_\alpha(\mathbf{w}) \leq H_\alpha(\mathbf{w}'), \quad (123)$$

for any quasi-distributions  $\mathbf{w}, \mathbf{w}'$  that obey  $\mathbf{w} \succ \mathbf{w}'$ .  $\square$

While we have integers  $a, b$  such that  $0 < \alpha = \frac{2a}{2b-1} < 1$  and  $H_\alpha(\mathbf{w})$  is well-defined, it turns out that monotonicity does not hold if we drop the condition  $a \geq b$ . If  $\alpha = 2a/(2b-1)$  with  $\alpha < 1$ , then  $g(w) := \frac{-w^{\alpha-1}}{\alpha-1}$  is no longer monotonic for all  $w \in \mathbb{R}$ , and the problem occurs when comparing  $w_i < 0$  and  $w_j > 0$ . As an example of this dependence of monotonicity on the domain of the function, consider the function  $g(x) = \frac{1}{x^3}$  which is monotone decreasing on both  $x < 0$  and  $x > 0$  separately, however it is not monotone on the full real-line.

We note that if  $\alpha = r/s$  with both  $r$  and  $s$  odd, then  $H_\alpha$  is not well-defined for all quasi-distributions, although if the actual set of quasi-distributions has a sufficiently bounded levels negativity, then  $\log \sum_{\mathbf{z}} W_\rho(\mathbf{z})^\alpha$  can still be obtained for  $r$  odd, provided the total sum is never negative. This occurs for Wigner distributions, where we have  $|W_\rho(\mathbf{z})| \leq 1/d$  for a  $d$ -dimensional quantum system.

We also note that the set  $F := \{2a/(2b-1) : a, b \in \mathbb{N} \text{ and } a \geq b\}$  is dense in the reals  $\mathbb{R}_{>1}$ , as any rational  $c/d$  with  $c > d$  can be approximated by  $c2^n/(d2^n-1) \in F$  for sufficiently large  $n$ .

We have that  $H_\alpha$  is additive on products of quasi-distributions, which we state for completeness.

**Proposition 21.** *For any  $\mathbf{w} \in \mathcal{Q}_N$  and  $\mathbf{w}' \in \mathcal{Q}_{N'}$ , we have*

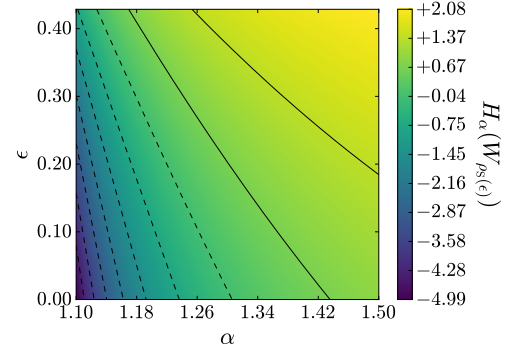
$$H_\alpha(\mathbf{w} \otimes \mathbf{w}') = H_\alpha(\mathbf{w}) + H_\alpha(\mathbf{w}'), \quad (124)$$

where  $\alpha = 2a/(2b-1)$  with positive integers  $a \geq b$ .

*Proof.*

$$\begin{aligned} H_\alpha(\mathbf{w} \otimes \mathbf{w}') &= \frac{1}{1-\alpha} \log \sum_{i,j} [w_i w'_j]^\alpha \\ &= \frac{1}{1-\alpha} \log \left[ \sum_i w_i^\alpha \sum_j w'_j{}^\alpha \right] \\ &= \frac{1}{1-\alpha} \log \left[ \sum_i w_i^\alpha \right] + \frac{1}{1-\alpha} \log \left[ \sum_j w'_j{}^\alpha \right] \\ &= H_\alpha(\mathbf{w}) + H_\alpha(\mathbf{w}'). \end{aligned} \quad (125)$$

$\square$



Supplementary Figure 2. **Rényi entropy  $H_\alpha$  of the  $\epsilon$ -noisy Strange state.**  $H_\alpha$  is increasing in  $\alpha$  and  $\epsilon$ . The rightmost dashed contour line corresponds to  $H_\alpha(W_{\rho_S(\epsilon)}) = 0$ . In particular,  $H_\alpha(W_{\rho_S(0)}) = 0$  occurs at  $\alpha \approx 1.31$ .

Applied to Wigner distributions  $W_\rho(\mathbf{z})$  for a quantum state  $\rho$  we have

$$H_\alpha(W_\rho) := \frac{1}{1-\alpha} \log \sum_{\mathbf{z}} W_\rho(\mathbf{z})^\alpha, \quad (126)$$

where  $\alpha$  can take values of the form  $2a/(2b-1)$ , for non-negative integers  $a, b$ . For the noisy Strange state  $\rho = \rho_S(\epsilon)$ , this becomes

$$H_\alpha(W_{\rho_S(\epsilon)}) = \frac{1}{1-\alpha} \log \left[ 8 \left( \frac{1}{6} - \frac{1}{18}\epsilon \right)^\alpha + \left( -\frac{1}{3} + \frac{4}{9}\epsilon \right)^\alpha \right] \quad (127)$$

We now show the following equivalence between Wigner negativity and negative Rényi entropies.

**Theorem 10.** *A quantum state  $\rho$  has Wigner negativity if and only if  $H_\alpha(W_\rho) < 0$  for some  $\alpha = \frac{2a}{2b-1}$ , with positive integers  $a \geq b$ .*

*Proof.* If we have  $H_\alpha(W_\rho) < 0$  for  $\alpha > 1$ , then it follows that  $\log \sum_{\mathbf{z}} W_\rho(\mathbf{z})^\alpha > 1$ . However, it is known that  $H_\alpha$  is always non-negative on probability distributions, so  $W_\rho(\mathbf{z})$  must be a strict quasi-distribution with negativity.

Conversely, suppose  $\rho$  has negativity in its Wigner representation. This in particular implies that  $\sum_{\mathbf{z}} |W_\rho(\mathbf{z})| > 1$ . We now consider  $\sum_{\mathbf{z}} W_\rho(\mathbf{z})^{\frac{2a}{2b-1}}$  for positive integers  $a \geq b$ . We have that

$$\begin{aligned} \sum_{\mathbf{z}} W_\rho(\mathbf{z})^{\frac{2a}{2b-1}} &= \sum_{\mathbf{z}} |W_\rho(\mathbf{z})|^{\frac{2a}{2b-1}} \\ &= \sum_{\mathbf{z}} |W_\rho(\mathbf{z})|^{1+\epsilon}, \end{aligned} \quad (128)$$

where  $\epsilon = \alpha - 1 = \frac{2a}{2b-1} - 1 > 0$ . By choosing the positive integers  $a$  and  $b$  sufficiently large we can make  $\epsilon$  arbitrarily close to 0. This implies that there exists a

sequence  $\epsilon_n = \alpha_n - 1 = \frac{2a_n}{2b_n-1} - 1$  with integer pairs  $a_n, b_n$  such that

$$\sum_{\mathbf{z}} |W_\rho(\mathbf{z})|^{1+\epsilon_n} \rightarrow \sum_{\mathbf{z}} |W_\rho(\mathbf{z})| > 1, \quad (129)$$

as  $n$  increases. Since  $\frac{1}{1-\alpha_n} < 0$  for any  $n$  it follows that  $H_{\alpha_n}(W_\rho) = \frac{1}{1-\alpha_n} \log \sum_{\mathbf{z}} W_\rho(\mathbf{z})^{\alpha_n} < 0$  at some point in the sequence, which completes the proof.  $\square$

Therefore Wigner negativity coincides with negativity of a Rényi entropy.

### Rényi divergences on quasi-distributions

We now define the  $\alpha$ -Rényi divergence for a quasi-distribution  $\mathbf{w} \in \mathcal{Q}_N$  relative to a full-rank probability distribution  $\mathbf{r} \in \mathcal{Q}_N^+$  as

$$D_\alpha(\mathbf{w}||\mathbf{r}) := \frac{1}{\alpha-1} \log \sum_i w_i^\alpha r_i^{1-\alpha}, \quad (130)$$

where  $\alpha = 2a/(2b-1)$  for positive integers  $a \geq b$ . We now have the following result that relates the Rényi divergence to the Rényi entropy on a dense subset of probability distributions.

**Proposition 22.** *Let  $\mathbf{w} \in \mathcal{Q}_N$  be a quasi-distribution and  $\mathbf{r} \in \mathcal{Q}_N^+$  a probability distribution with positive rational entries given by  $r_i = a_i/K$  for positive integers  $a_i$  and  $K = \sum_{i=1}^N a_i$ . Then,*

$$H_\alpha(\Gamma_\alpha(\mathbf{w})) = K - D_\alpha(\mathbf{w}||\mathbf{r}), \quad (131)$$

where  $\alpha = 2a/(2b-1)$  for positive integers  $a \geq b$ .

*Proof.* For the given form of  $\alpha$  and positive  $r_i$ ,  $i = 1, \dots, N$ ,  $D_\alpha(\mathbf{w}||\mathbf{r})$  is well-defined for all  $\mathbf{w}$ . From the definition of  $\Gamma_\alpha$  we have

$$\Gamma_\alpha(\mathbf{w}) = \bigoplus_{i=1}^N w_i(1/a_i, 1/a_i, \dots, 1/a_i). \quad (132)$$

This directly leads to

$$\begin{aligned} H_\alpha(\Gamma_\alpha(\mathbf{w})) &= \frac{1}{1-\alpha} \log \sum_{i=1}^N a_i \left( \frac{w_i}{a_i} \right)^\alpha \\ &= \frac{1}{1-\alpha} \log \sum_{i=1}^N w_i^\alpha a_i^{1-\alpha} \\ &= \frac{1}{1-\alpha} \log K^{1-\alpha} \sum_{i=1}^N w_i^\alpha r_i^{1-\alpha} \\ &= K + \frac{1}{1-\alpha} \log \sum_{i=1}^N w_i^\alpha r_i^{1-\alpha} \\ &= K - D_\alpha(\mathbf{w}||\mathbf{r}). \end{aligned} \quad (133)$$

$\square$

With this we now establish monotonicity for Rényi relative divergences.

**Proposition 23.** *Let  $\alpha = \frac{2a}{2b-1}$  for any positive integers  $a, b$  with  $a \geq b$ . Let  $\mathbf{w} \in \mathcal{Q}_N$ ,  $\mathbf{w}' \in \mathcal{Q}_{N'}$  and  $\mathbf{r} \in \mathcal{Q}_N^+$ ,  $\mathbf{r}' \in \mathcal{Q}_{N'}^+$ , with positive rational components  $a_i/K$  and  $a'_i/K$  respectively. Then  $D_\alpha(\mathbf{w}||\mathbf{r}) \geq D_\alpha(\mathbf{w}'||\mathbf{r}')$ , whenever  $(\mathbf{w}, \mathbf{r}) \succ (\mathbf{w}', \mathbf{r}')$ .*

*Proof.* The statement  $(\mathbf{w}, \mathbf{r}) \succ (\mathbf{w}', \mathbf{r}')$  is equivalent to  $\Gamma_\alpha(\mathbf{w}) \succ \Gamma_\alpha(\mathbf{w}')$ . Therefore, due to Theorem 9 we have

$$H_\alpha(\Gamma_\alpha(\mathbf{w})) \leq H_\alpha(\Gamma_\alpha(\mathbf{w}')) \quad (134)$$

which, due to Proposition 22, is equivalent to

$$D_\alpha(\mathbf{w}||\mathbf{r}) \geq D_\alpha(\mathbf{w}'||\mathbf{r}'). \quad (135)$$

$\square$

Since the rationals are dense in the real numbers we can assume any Wigner distribution considered has rational components without affecting results. We now specialize to Wigner distributions of quantum states, and quantum operations that map the set of free states  $\mathcal{F}$  into itself.

**Theorem 11.** *Let  $\tau$  be in the interior of  $\mathcal{F}$ . If  $\alpha = \frac{2a}{2b-1}$  for positive integers  $a, b$  with  $a \geq b$ , then the  $\alpha$ -Rényi divergence  $D_\alpha(W_\rho||W_\tau)$  is well-defined for all states  $\rho$ , and the following hold:*

1.  $D_\alpha(W_\rho||W_\tau) \geq 0$  for all quantum states  $\rho$ .
2.  $D_\alpha(W_\rho||W_\tau) = 0$  if and only if  $\rho = \tau$ .
3.  $D_\alpha(W_{\rho^{\otimes n}}||W_{\tau^{\otimes n}}) = nD_\alpha(W_\rho||W_\tau)$  for any  $n \in \mathbb{N}$ .
4.  $D_\alpha(W_\rho||W_\tau) \geq D_\alpha(W_{\mathcal{E}(\rho)}||W_{\mathcal{E}(\tau)})$  for any free operation  $\mathcal{E}$  such that  $\mathcal{E}(\tau)$  is in the interior of  $\mathcal{F}$ .

*Proof.* Since  $\tau$  is in the interior of  $\mathcal{F}$ , its Wigner function obeys  $W_\tau(\mathbf{z}) > 0$  for all points  $\mathbf{z}$  in the phase space. In general,  $W_\rho(\mathbf{z})$  is a quasi-distribution, but for  $\alpha = 2a/2b-1$  we have that  $W_\rho(\mathbf{z})^\alpha \geq 0$ . Therefore  $D_\alpha(W_\rho||W_\tau)$  is always well-defined and real-valued.

1.  $D_\alpha$  is Schur-convex and every pair  $(W_\rho, W_\tau)$  majorizes the pair  $(W_{1/d}, W_{1/d})$ , so

$$\begin{aligned} D_\alpha(W_\rho||W_\tau) &\geq D_\alpha(W_{1/d}||W_{1/d}) = \\ &= \frac{1}{\alpha-1} \log \sum_{\mathbf{z}} W_{1/d}(\mathbf{z}) = 0. \end{aligned} \quad (136)$$

2. In the inequality of property 1, equality holds iff  $L_{\rho|\tau}(x) = L_{1/d|1/d}(x) = L_{\tau|\tau}(x)$  for all  $x \in [0, 1]$  due to Proposition 9 which in turn holds iff  $\rho = \tau$ .

3. This property follows from the multiplicativity of the Wigner distribution. In particular,

$$\begin{aligned}
D_\alpha(W_{\rho^{\otimes n}} || W_{\tau^{\otimes n}}) &= \\
\frac{1}{\alpha-1} \log \sum_{\mathbf{z}} W_{\rho^{\otimes n}}(\mathbf{z})^\alpha W_{\tau^{\otimes n}}(\mathbf{z})^{1-\alpha} &= \\
\frac{1}{\alpha-1} \log \sum_{\mathbf{z}} \prod_{i=1}^n W_{\rho}(z_i)^\alpha W_{\tau}(z_i)^{1-\alpha} &= \\
\frac{1}{\alpha-1} \log \prod_{i=1}^n \sum_{z_i} W_{\rho}(z_i)^\alpha W_{\tau}(z_i)^{1-\alpha} &= \\
\sum_{i=1}^n \frac{1}{\alpha-1} \log \sum_{z_i} W_{\rho}(z_i)^\alpha W_{\tau}(z_i)^{1-\alpha} &= \\
= n D_\alpha(W_\rho || W_\tau). & \quad (137)
\end{aligned}$$

4. This follows immediately from the fact that  $(W_\rho, W_\tau) \succ (W_{\mathcal{E}(\rho)}, W_{\mathcal{E}(\tau)})$  for any free quantum channel  $\mathcal{E}$  that sends  $\tau$  into the interior of  $\mathcal{F}$ , and the Schur-convexity of  $D_\alpha$ .  $\square$

We are now in a position to derive general magic state distillation bounds.

**Theorem 12.** *Consider a general magic state distillation protocol on odd prime dimension qudits, that converts a magic state  $\rho^{\otimes n} \rightarrow \mathcal{E}(\rho^{\otimes n}) = \rho'^{\otimes m}$  and let  $\tau$  be any full-rank stabilizer reference state on a qudit. Then,*

*the distillation rate  $R := m/n$  is upper bounded as*

$$R \leq R_\alpha := \frac{D_\alpha(W_\rho || W_\tau)}{\tilde{D}_\alpha(\rho', \tau')}, \quad (138)$$

*where  $\alpha = \frac{2a}{2b-1}$  for any positive integers  $a, b$  with  $a \geq b$  and the average divergence per qudit*

$$\tilde{D}_\alpha(\rho', \tau') := \frac{1}{m} D_\alpha(W_{\rho'^{\otimes m}} || W_{\tau_m'}), \quad (139)$$

*between the output magic state  $\rho'^{\otimes m}$  and  $\tau_m' = \mathcal{E}(\tau^{\otimes n})$ .*

*Proof.* The bound is a direct consequence of the properties of the  $\alpha$ -Rényi divergence in Theorem 11.

Due to the action of the magic protocol channel, we get

$$D_\alpha(W_{\rho^{\otimes n}} || W_{\tau^{\otimes n}}) \geq D_\alpha(W_{\rho'^{\otimes m}} || W_{\tau_m'}). \quad (140)$$

We can use the additivity to rewrite this as

$$n D_\alpha(W_\rho || W_\tau) \geq m \frac{1}{m} D_\alpha(W_{\rho'^{\otimes m}} || W_{\tau_m'}). \quad (141)$$

Since  $\rho'^{\otimes m} \neq \tau_m'$ , we have  $D_\alpha(W_{\rho'^{\otimes m}} || W_{\tau_m'}) > 0$ , which directly leads to the bound

$$\frac{m}{n} \leq \frac{D_\alpha(W_\rho || W_\tau)}{\tilde{D}_\alpha(\rho', \tau')}. \quad (142)$$

$\square$

## SUPPLEMENTARY REFERENCES

- [1] Veitch, V., Ferrie, C., Gross, D. & Emerson, J. Negative quasi-probability as a resource for quantum computation. *New J. Phys.* **14**, 113011 (2012).
- [2] Vourdas, A. Quantum systems with finite hilbert space. *Rep. Prog. Phys.* **67**, 267–320 (2004).
- [3] Gross, D. Hudson’s theorem for finite-dimensional quantum systems. *J. Math. Phys.* **47**, 122107 (2006).
- [4] Wang, X., Wilde, M. M. & Su, Y. Quantifying the magic of quantum channels. *New J. Phys.* **21**, 103002 (2019).
- [5] Mirsky, L. On the trace of matrix products. *Mathematische Nachrichten* **20**, 171–174 (1959).
- [6] Horodecki, M. & Oppenheim, J. Fundamental limitations for quantum and nanoscale thermodynamics. *Nat. Commun.* **4**, 2059 (2013).
- [7] Brandão, F., Horodecki, M., Ng, N., Oppenheim, J. & Wehner, S. The second laws of quantum thermodynamics. *PNAS* **112**, 3275–3279 (2015).
- [8] Lostaglio, M. An introductory review of the resource theory approach to thermodynamics. *Rep. Prog. Phys.* **82**, 114001 (2019).
- [9] Marshall, A. W., Olkin, I. & Arnold, B. C. *Inequalities: Theory of Majorization and Its Applications* (Springer, 2011).
- [10] Ruch, E., Schraner, R. & Seligman, T. H. The mixing distance. *J. Chem. Phys.* **69**, 386–392 (1978).
- [11] Renes, J. M. Relative submajorization and its use in quantum resource theories. *J. Math. Phys.* **57**, 122202 (2016).
- [12] Buscemi, F. & Gour, G. Quantum relative lorenz curves. *Phys. Rev. A* **95**, 012110 (2017).
- [13] Ash, R. *Information Theory* (Dover Publications Inc., 1965).
- [14] Schur, I. Über eine klasse von mittelbildungen mit anwendungen auf die determinantentheorie. *Berliner Mathematische Gesellschaft* **22**, 9–20 (1923).
- [15] Ostrowski, A. M. Sur quelques applications des fonctions convexes et concaves au sens de i. schur. *Journal de Mathématiques Pures et Appliquées* **31**, 253–292 (1952).
- [16] Manfredi, G. & Feix, M. R. Entropy and Wigner functions. *Phys. Rev. E* **62**, 4665–4674 (2000).
- [17] Gnutzmann, S. & Życzkowski, K. Rényi-wehrl entropies as measures of localization in phase space. *J. Phys. A Math.* **34**, 10123–10139 (2001).