Mixed-state entanglement from local randomized measurements
–
Supplemental Material

Andreas Elben,[1,2,*] Richard Kueng,[3,*] Hsin-Yuan (Robert) Huang,[4,5] Rick van Bijnen,[1,2] Christian Kokail,[1,2] Marcello Dalmonte,[6,7] Pasquale Calabrese,[6,7,8] Barbara Kraus,[9] John Preskill,[4,5,10,11] Peter Zoller,[1,2] and Benoît Vermersch[1,2,12]

[1] *Center for Quantum Physics, University of Innsbruck, Innsbruck A-6020, Austria*
[2] *Institute for Quantum Optics and Quantum Information of the Austrian Academy of Sciences, Innsbruck A-6020, Austria*
[3] *Institute for Integrated Circuits, Johannes Kepler University Linz, Altenbergerstrasse 69, 4040 Linz, Austria*
[4] *Institute for Quantum Information and Matter, Caltech, Pasadena, CA, USA*
[5] *Department of Computing and Mathematical Sciences, Caltech, Pasadena, CA, USA*
[6] *The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy*
[7] *SISSA, via Bonomea 265, 34136 Trieste, Italy*
[8] *INFN, via Bonomea 265, 34136 Trieste, Italy*
[9] *Institute for Theoretical Physics, University of Innsbruck, A–6020 Innsbruck, Austria*
[10] *Walter Burke Institute for Theoretical Physics, Caltech, Pasadena, CA, USA*
[11] *AWS Center for Quantum Computing, Pasadena, CA, USA*
[12] *Univ. Grenoble Alpes, CNRS, LPMMC, 38000 Grenoble, France*

Appendix A: The p_3-PPT condition

In this section we present, prove and discuss the p_3-PPT condition. The p_3-PPT condition is the contrapositive of the following statement about moments of positive semidefinite matrices with unit trace.

Proposition 1. *For every positive semidefinite matrix X with unit trace ($\mathrm{Tr}(X) = 1$) it holds that*

$$\mathrm{tr}(X^2)^2 \leq \mathrm{tr}(X^3). \tag{A1}$$

Note that Eq. (A1) resembles the following well-known monotonicity relation among Rényi entropies (see e.g., Ref. [1]):

$$S_3(\rho) \leq S_2(\rho) \;\text{ for }\; S_n(\rho) = \tfrac{1}{1-n}\log_2\big(\mathrm{tr}(\rho^n)\big). \tag{A2}$$

However, this relation only applies to density matrices, i.e. positive semidefinite matrices with unit trace. The p_3-PPT condition, in contrast, is designed to test the absence of positive semidefiniteness. Hence, it is crucial to have a condition that does not break down if the matrix in question has negative eigenvalues. Rel. (A1) (and its direct proof provided in the next subsection) do achieve this goal, while an argument based on monotonicity relations between Rényi entropies can break down, because the logarithm of non-positive numbers is not properly defined.

1. Proof of the p_3-PPT condition

Let X be a Hermitian $d \times d$ matrix with eigenvalue decomposition $X = \sum_{i=1}^d \lambda_i |x_i\rangle\langle x_i|$. For $p \geq 1$, we introduce the Schatten-p norms

$$\|X\|_p = \Big(\sum_{i=1}^d |\lambda_i|^p\Big)^{1/p} = \mathrm{Tr}\left(|X|^p\right)^{1/p},$$

where $|X| = \sqrt{X^2} = \sum_{i=1}^d |\lambda_i||x_i\rangle\langle x_i|$ denotes the (matrix-valued) absolute value. The Schatten-p norms encompass most widely used matrix norms in quantum information. Concrete examples are the trace norm ($p = 1$), the Hilbert-Schmidt/Frobenius norm ($p = 2$) and the operator/spectral norm ($p = \infty$). Each Schatten-p norm corresponds to the usual vector ℓ_p-norm of the vector of eigenvalues $\lambda = (\lambda_1, \ldots, \lambda_d)^T \in \mathbb{R}^d$:

$$\|\lambda\|_{\ell_p} = \Big(\sum_{i=1}^d |\lambda_i|^p\Big)^{1/p} \quad \text{for } p \geq 1. \tag{A3}$$

Hence, Schatten-p norms inherit many desirable properties from their vector-norm counterparts. Here, we shall use vector norm relations to derive a relation among Schatten-p norms. It is based on Hoelder's inequality that relates the inner product

$$\langle v, w\rangle = \sum_{i=1}^d v_i w_i \quad \text{for } v, w \in \mathbb{R}^d \tag{A4}$$

to a combination of ℓ_p norms.

Fact 1 (Hoelder's inequality for vector norms)**.** *Fix $p, q \geq 1$ such that $1/p + 1/q = 1$. Then,*

$$|\langle v, w\rangle| \leq \sum_{i=1}^d |v_i w_i| \leq \|v\|_{\ell_p}\|w\|_{\ell_q} \tag{A5}$$

for any $v, w \in \mathbb{R}^d$.

[*] These authors contributed equally.

The well-known Cauchy-Schwarz inequality is a special case of this fact. Set $p = q = 1/2$ to conclude

$$|\langle v, w\rangle| \leq \|v\|_{\ell_2}\|w\|_{\ell_2} = \langle v, v\rangle^{1/2}\langle w, w\rangle^{1/2}. \tag{A6}$$

At the heart of our proof for the p_3-PPT condition is a simple relation between Schatten-p norms of orders $p = 1, 2, 3$.

Lemma 1. *The following norm relation holds for every Hermitian matrix X:*

$$\|X\|_2^4 \leq \|X\|_1\|X\|_3^3$$

Proof. Let $\lambda = (\lambda_1, \ldots, \lambda_d)^T$ be the d-dimensional vector of eigenvalues of X. Apply Hoelder's inequality with $p = 3, q = 3/2$ to the inner product of this vector of eigenvalues with itself:

$$\mathrm{Tr}(X^2) = \langle \lambda, \lambda\rangle \leq \|\lambda\|_{\ell_3}\|\lambda\|_{\ell_{2/3}} = \|X\|_3\|\lambda\|_{\ell_{3/2}}. \tag{A7}$$

Next, we apply Cauchy-Schwarz to the remaining $\ell_{3/2}$-norm:

$$\begin{aligned}\|\lambda\|_{\ell_{3/2}} =& \Big(\sum_{i=1}^d |\lambda_i|^{3/2}\Big)^{2/3} = \Big(\sum_{i=1}^d |\lambda_i||\lambda_i|^{1/2}\Big)^{2/3}\\ \leq& \Big(\Big(\sum_{i=1}^d |\lambda_i|^2\Big)^{1/2}\Big(\sum_{i=1}^d |\lambda_i|^{2/2}\Big)^{1/2}\Big)^{2/3}\\ =& \|\lambda\|_{\ell_2}^{2/3}\|\lambda\|_{\ell_1}^{1/3} = \|X\|_2^{2/3}\|X\|_1^{1/3}.\end{aligned}$$

Inserting this relation into Eq. (A7) reveals

$$\|X\|_2^2 \leq \|X\|_2^{2/3}\|X\|_1^{1/3}\|X\|_3$$

which is equivalent to the claim (take the 3rd power and divide by $\|X\|_2^2$). □

Proposition 1 is an immediate consequence of Lemma 1 and elementary properties of positive semidefinite matrices. Recall that a Hermitian $d \times d$ matrix is positive semidefinite (psd) if every eigenvalue is nonnegative. This in turn ensures $|X| = X$ and, by extension, $\|X\|_p = \mathrm{Tr}(X^p)^{1/p}$ for all $p \geq 1$.

2. Discussion and potential generalizations

The p_3-PPT condition tests the absence of positive semidefiniteness based on moments $\mathrm{Tr}(X^p)$ of order $p = 1, 2, 3$. It is natural to wonder whether higher order moments allow the construction of more refined tests. It is possible to show that every positive semidefinite matrix X with unit trace must obey

$$\mathrm{tr}(X^{p-1})^{p-1} \leq \mathrm{tr}(X^p)^{p-2} \quad \text{for all } p > 2. \tag{A8}$$

As this is a direct extension of the p_3-PPT condition ($p = 3$), we omit the proof. Unfortunately, we found numerically that these direct extensions actually produce *weaker* tests for the absence of positive semidefiniteness, i.e. there exist matrices X that violate the p_3-PPT condition but satisfy Rel. (A8) for higher moments $p \geq 4$. This is not completely surprising, since Rel. (A8) compares (powers of) neighboring matrix moments with order $(p-1)$ and p. As p increases, these matrix moments suppress contributions of small eigenvalues ever more strongly. In the case of partially transposed quantum states, the eigenvalues are required to sum up to one and must be contained in the interval $[-1/2, 1]$ [2]. Thus, the negative eigenvalues can never dominate the spectrum and high matrix moment tests for the existence of negative eigenvalues suffer from suppression effects.

This observation suggests that powerful tests for negative eigenvalues should involve *all* matrix moments $\mathrm{tr}(X^p)$ up to a certain order $p_{\max}$. It is useful to change perspective in order to reason about potential improvments. The p_3-PPT condition checks whether the following inequality is true:

$$F_3(X) = -\mathrm{tr}(X^3) + \mathrm{tr}(X^2)^2 > 0. \tag{A9}$$

For matrices X with unit trace, we can reinterpret the matrix-valued function $F_3(X)$ as a sum of (identical) degree-3 polynomials applied to all eigenvalues $\lambda_1, \ldots, \lambda_d$ of X. Set $p_2 = \mathrm{tr}(X^2)$ and use $\mathrm{tr}(X) = \sum_{i=1}^d \lambda_i = 1$ to conclude

$$\begin{aligned}F_3(X) =& -\mathrm{tr}(X^3) + 2p_2\mathrm{tr}(X^2) - p_2^2\mathrm{tr}(X)\\ =& \sum_{i=1}^d \left(-\lambda_i^3 + 2p_2\lambda_i^2 - p_2^2\lambda_i\right)\\ =& \sum_{i=1}^d -\lambda_i(\lambda_i - p_2)^2 =: \sum_{i=1}^d f_3(\lambda_i).\end{aligned} \tag{A10}$$

Note that the polynomial

$$f_3(x) = -x(x - p_2)^2 \quad \text{for} \quad x \in \mathbb{R} \tag{A11}$$

depends on p_2 and, by extension, also on the matrix X. We will come back to this aspect later. For now, we point out that – regardless of the actual value of p_2 – this polynomial has three interesting properties:

$$\begin{array}{ll} f_3(x) \leq 0 & \text{if } x > 0,\\ f_3(0) = 0, & \\ f_3(x) > 0 & \text{if } x < 0.\end{array} \tag{A12}$$

These properties reflect the behavior of another well-known function – the *(negated) rectifier function* (ReLU):

$$r(-x) = \max\{0, -x\} = \begin{cases} 0 & \text{if } x \geq 0,\\ |x| & \text{if } x < 0.\end{cases} \tag{A13}$$

See Figure A1 for a visual comparison. Applying the (negated) rectifier function to the eigenvalues of X would recover the negativity:

$$\mathcal{N}(X) = \sum_{\lambda_i < 0} |\lambda_i| = \sum_{i=1}^d r(-\lambda_i). \tag{A14}$$

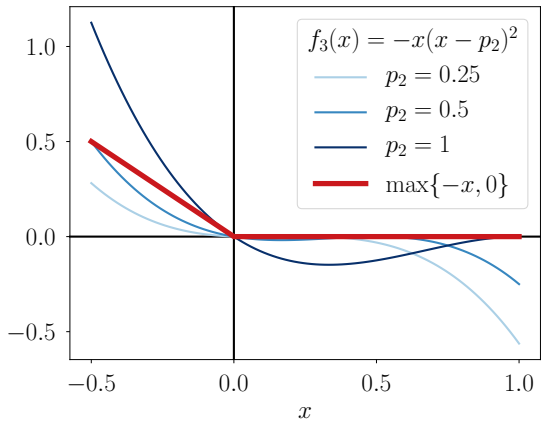


FIG. A1. Comparison of $f_3(x) = -x(x-p_2)^2$ with the negated rectifier function $r(-x) = \max\{-x, 0\}$ for different values of p_2 in the relevant interval $[-1/2, 1]$ [2].

Hence, it is instructive to interpret $F_3(X)$ as a polynomial approximation to the (non-analytic) negativity function.

On the level of polynomials, the condition $f_3(x) \leq 0$ whenever $x > 0$ is most important. It implies that positive eigenvalues of X can never increase the value of $F_3(X) = \sum_{i=1}^{d} f_3(\lambda_i)$. In particular, $F_3(X) \leq 0$ whenever X is positive semidefinite – as stated in Proposition 1. The p_3-PPT condition is sound, i.e. it has no false positives.

Conversely, $f_3(x) > 0$ for $x < 0$ implies that $F_3(X)$ can become positive if X has negative eigenvalues. Hence, the p_3-PPT condition is not vacuous. It is capable of detecting negative eigenvalues in many, but not all, unit-trace matrices X.

Let us now return to the (matrix-dependent) parameter choice in Eq. (A11). In principle, every polynomial of the form $f_3^{(a)}(x) = -x(x-a)^2$ with $a \in \mathbb{R}$ obeys the important structure constraints (A12) and therefore produces a sound test for negative eigenvalues. For fixed X, the associated matrix polynomial evaluates to

$$F_3^{(a)}(X) = -\mathrm{tr}(X^3) + 2a\mathrm{tr}(X^2) - a^2\mathrm{tr}(X). \tag{A15}$$

We can optimize this expression over the parameter $a \in \mathbb{R}$ to make the test as strong as possible. The optimal choice is $a_\sharp = \mathrm{tr}(X^2)/\mathrm{tr}(X)$ and produces a matrix polynomial that obeys $F_3^{(a^\sharp)}(X) \geq \max_{a\in\mathbb{R}} F_3^{(a)}(X)$ for X fixed. If X has also unit trace, the optimal parameter becomes $a_\sharp = p_2$ and produces the p_3-PPT condition (A9).

This construction of PPT conditions readily extends to higher order polynomials $f_p(x) = a_p x^p + \cdots a_1 x + a_0$. Increasing the degree p produces more expressive ansatz functions that can approximate the (negated) rectifier function – and its core properties – ever more accurately. Viewed from this angle, it becomes apparent that measuring more matrix moments can produce stronger tests for detecting negative eigenvalues. However, it is not so obvious how to choose the parameters $a_p, \ldots, a_0$ "optimally", or what "optimally" actually means in this context. Some well-known polynomial approximations of the rectifier function $r(-x)$ – like Taylor expansions of $s(-x) = \ln(1 + \mathrm{e}^{-x})$ (the "softplus" function) – are not well-suited for this task, because $s(-x) > 0$ even for $x > 0$. This in turn would imply that the associated test condition may not be sound. We believe that a thorough analysis of these questions is timely and interesting, but would go beyond the scope of this work. We intend to address it in future research.

Appendix B: p_3-PPT condition for Werner States

Werner states are bipartite quantum states in a Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ with dimensions $d_A = d_B \equiv d$, defined as

$$\rho_W = \alpha \binom{d+1}{2}^{-1} \Pi_+ + (1-\alpha)\binom{d}{2}^{-1} \Pi_- \tag{B1}$$

with parameter $\alpha \in [0, 1]$ and $\Pi_\pm = \frac{1}{2}(\mathbb{I} \pm \Pi_{12})$ projectors onto symmetric $\mathcal{H}_+$ and anti-symmetric $\mathcal{H}_-$ subspaces of $\mathcal{H} = \mathcal{H}_+ \oplus \mathcal{H}_-$, respectively [3]. Here, $\Pi_{12} = \sum_{i,j=1}^{d} |i\rangle\langle j| \otimes |j\rangle\langle i|$ is the swap operator. We note that the eigenvalues of ρ_W are thus given as $\lambda_+ = \alpha\binom{d+1}{2}^{-1}$ with multiplicity $\binom{d+1}{2}$ and $\lambda_- = (1-\alpha)\binom{d}{2}^{-1}$ with multiplicity $\binom{d}{2}$. The reduced state ρ_A of qudit A is given by $\rho_A = \mathrm{Tr}_B[\rho_W] = \mathbb{I}_A/d$.

Using furthermore that $\Pi_\pm^{T_A} = 1/2(\Delta_1 \pm (d \pm 1)\Delta_0)$ with $\Delta_0 = |\phi_+\rangle\langle\phi_+|$ being a projector onto the maximally entangled state and $\Delta_1 = \mathbb{I} - \Delta_0$ [3], we find

$$\rho_W^{T_A} = \frac{2\alpha - 1}{d}\Delta_0 + \frac{1 + d - 2\alpha}{d}\frac{\Delta_1}{d^2 - 1} \tag{B2}$$

with eigenvalues $\lambda_0 = (2\alpha - 1)/d$ with multiplicity 1 and $\lambda_1 = (1 + d - 2\alpha)/d(d^2 - 1)$ with multiplicity $d^2 - 1$.

We note that, for any d, $\lambda_0 < 0$ for $0 \leq \alpha < 1/2$. Thus, using the PPT condition, we find that ρ_W is entangled for $0 \leq \alpha < 1/2$. Using the explicit expression of the eigenvalues, we can furthermore determine $\mathrm{Tr}\left[(\rho^{T_A})^n\right]$ for any n. We find for all local dimensions d

$$\mathrm{Tr}\left[(\rho^{T_A})^2\right]^2 > \mathrm{Tr}\left[(\rho^{T_A})^3\right] \quad \text{for} \quad 0 \leq \alpha < \frac{1}{2} \tag{B3}$$

Thus, for Werner states the p_3-PPT condition is equivalent to the full PPT condition. It can be furthermore shown that Werner states are separable for $\alpha \geq 1/2$ [3]. Thus, for Werner states, the p_3-PPT condition is a necessary and sufficient condition for bipartite entanglement. This also holds true for "isotropic" states of the form $\rho = \alpha\mathbb{1}/d^2 + (1-\alpha)|\phi_+\rangle\langle\phi_+|$, which are closely related.

We note that Werner states can have non-positive PT-moments. For local dimension $d > 3$ there exists a parameter interval $[0, \alpha^*)$ such that the associated Werner

state (B1) obeys $p_3 = \mathrm{Tr}\left[(\rho_W^{T_A})^3\right] < 0$ for all $\alpha \in [0, \alpha^*)$. This highlights that the logarithm of PT-moments, appearing also in the ratio $R_3 = -\log_2(p_3/\mathrm{Tr}[\rho^3])$, need not be properly defined, justifying a claim from the previous subsection. It is difficult to use entropic arguments for reasoning about relations between (logarithmic) PT-moments.

Finally, as shown in Ref. [4], we remark that R_3 is not an entanglement monotone. For separable Werner states with $1/2 \leq \alpha < 1/2 + 1/(2d)$, it holds that $0 < p_3 < \mathrm{Tr}[\rho^3]$. Thus, $R_3 = -\log_2(p_3/\mathrm{Tr}[\rho^3])$ can be greater than zero, even for separable states. Since R_3 equals zero for all product states, it is not an entanglement monotone [5].

Appendix C: Comparison of entanglement conditions for quench dynamics

In this section, we compare the diagnostic power of the full PPT-condition, the p_3-PPT condition and a condition based on purities of nested subsystems to detect bipartite entanglement of mixed states. Specifically, given a reduced density matrix ρ_{AB} in a bipartite system AB, we consider:

1. the PPT-condition detecting bipartite entanglement between A and B for a strictly positive negativity $\mathcal{N}(\rho_{AB}) = \sum_{\lambda<0} |\lambda| > 0$, with λ the spectrum of $\rho_{AB}^{T_A}$ [5].

2. the p_3-PPT condition detecting bipartite entanglement between A and B for $1 - p_3/p_2^2 > 0$.

3. a condition based on the purity of nested subsystems detecting bipartite entanglement between A and B for $\mathrm{Tr}[\rho_A^2] < \mathrm{Tr}[\rho_{AB}^2]$ with $\rho_A = \mathrm{Tr}_B[\rho_{AB}]$ the reduced density matrix of subsystem A [5].

The latter 'purity' condition was used in previous experimental works measuring the second Rényi entropy [6–9] to reveal bipartite entanglement of weakly mixed states.

To test these conditions, we consider here, as an example, quantum states generated via quench dynamics in interacting spin models. Specifically, we study quenches in the XY-model with long-range interactions, as defined in Eq. (6) of the main text, in a total system with $N = 10$ spins. The initial separable product state is a Néel state $|\uparrow\downarrow\uparrow\downarrow\ldots\rangle$.

As shown in Fig. C1, the negativity (red lines) detects bipartite entanglement for all partitions sizes and all times after the quench. The p_3-PPT condition (blue lines) performs similar for the partitions considered in panel (b) and (c) and is thus able to detect bipartite entanglement for highly mixed states ρ_{AB} whose purity decreases to 0.3 for the panel (b) at late times. The p_3-PPT conditions fails however to detect the entanglement for the close-to completely mixed states of small partitions $|AB| = 4$ at late times, displayed in panel (a).

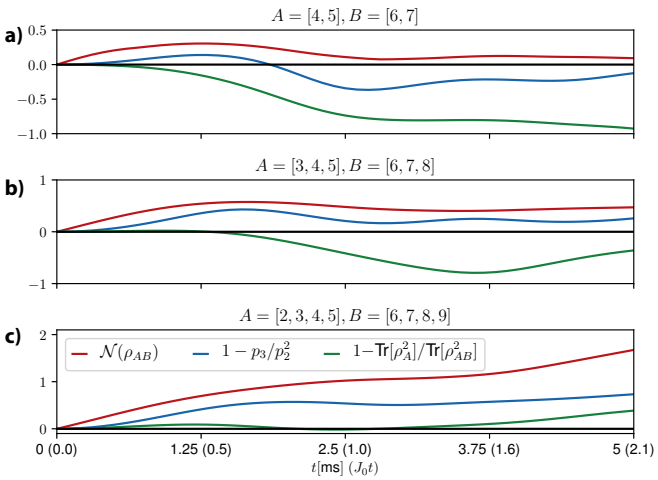


FIG. C1. Comparing conditions for bipartite entanglement between two subsystems A and B for states generated with quench dynamics governed by H_{XY} arising from an initial Néel state in a total system with $N = 10$ spins. Modeling the experiment of Ref. [9], we chose $J_0 = 420s^{-1}, \alpha = 1.24$, while other parameter choices lead to similar results. In all panels, and for all quantities, a strictly positive value signals bipartite entanglement.

This can be attributed to the fact that the p_3-PPT condition only relies on low order PT-moments. The purity condition (green lines) is only useful for the detection of entanglement for large partitions AB with $|AB| = 8$ (panel (c)). These remain weakly mixed during the entire time evolution, since the total system of $N = 10$ spins is described here by a pure state.

Appendix D: Error bars for PT moment predictions

Let us first review the data acquisition procedure. To obtain meaningful information about an N-qubit state ρ, we first perform a collection of random single qubit rotations: $\rho \mapsto u\rho u^\dagger$, where $u = u_1 \otimes \cdots \otimes u_N$ and each u_i is chosen from a unitary 3-design. Subsequently, we perform computational basis measurements and store the outcome:

$$\rho \mapsto u\rho u^\dagger \mapsto |k_1, \ldots, k_N\rangle. \tag{D1}$$

Here, $k_1, \ldots, k_N \in \{0, 1\}$ denote the measurement outcomes on qubits $1, \ldots, N$. As shown in [10–12], the outcome of this measurement provides a (single-shot) estimate for the unknown state:

$$\hat{\rho} = \bigotimes_{i=1}^{N} \left[3(u_i)^\dagger |k_i\rangle \langle k_i| u_i - \mathbb{I}_2\right] \tag{D2}$$

This tensor product is a random matrix – the unitaries $u^{(i)}$, as well as the observed outcomes k_i are random – that produces the true underlying state in expectation:

$$\mathbb{E}\left[\hat{\rho}\right] = \rho. \tag{D3}$$

Thus, the result of a (randomly selected) single-shot measurement provides a classical snapshot (D2) that reproduces the true underlying state in expectation. This desirable feature extends to density matrices of subsystems. Let $AB \subset \{1,\dots,N\}$ be a subset of $|AB| \leq N$ qubits and let $\rho_{AB} = \mathrm{Tr}_{\neg AB}(\rho)$ the associated reduced density matrix. Then,

$$\hat{\rho}_{AB} = \mathrm{Tr}_{\neg AB}(\hat{\rho}) = \bigotimes_{i\in AB} \left[3(u_i)^\dagger |k_i\rangle \langle k_i| u_i - \mathbb{I}_2\right] \quad \text{(D4)}$$
$$\text{obeys} \quad \mathbb{E}\left[\hat{\rho}_{AB}\right] = \rho_{AB}.$$

This feature can be used to estimate linear properties of the subsystem in question: $o = \mathrm{Tr}\left(O\rho_{AB}\right)$. Perform M independent repetitions of the data acquisition procedure and use them to create a collection of (independent) snapshots $\hat{\rho}_{AB}^{(1)},\dots,\hat{\rho}_{AB}^{(M)}$ – a "classical shadow" [12] – and form the empirical average of subsystem properties:

$$\hat{o} = \tfrac{1}{M}\sum_{r=1}^{M} \mathrm{Tr}\left(O\hat{\rho}_{AB}^{(r)}\right). \quad \text{(D5)}$$

Convergence to the target value $o = \mathbb{E}\left[\hat{o}\right] = \mathrm{Tr}(O\rho_{AB})$ is controlled by the variance. Chebyshev's inequality asserts

$$\Pr\left[|\hat{o} - o| \geq \epsilon\right] \leq \frac{\mathrm{Var}\left[\hat{o}\right]}{\epsilon^2} = \frac{\mathrm{Var}\left[\mathrm{Tr}(O_{AB}\hat{\rho}_{AB})\right]}{M\epsilon^2}. \quad \text{(D6)}$$

The remaining (single-shot) variance obeys the following useful relation.

Fact 2 (Proposition 3 in [12]). *Fix a subsystem AB and a linear function $\mathrm{Tr}(O\rho_{AB})$. Then, the single-shot variance associated with $\hat{\rho}_{AB}$ defined in Eq. (D4) obeys*

$$\mathrm{Var}\left[\mathrm{Tr}\left(O\hat{\rho}_{AB}\right)\right] \leq 2^{|AB|}\mathrm{Tr}\left(O^2\right). \quad \text{(D7)}$$

This inequality is true for any underlying state ρ and bounds the variance in terms of the subsystem dimension $d_A = 2^{|AB|}$ and the Hilbert-Schmidt norm (squared) of the observable O. Thus, roughly $M \approx 2^{|AB|}\mathrm{Tr}(O^2)/\epsilon^2$ measurement repetitions are necessary to predict o up to accuracy ϵ.

1. Predicting quadratic properties (p_2)

The formalism introduced above readily extends to predictions of higher order polynomials. The special case of quadratic functions has already been addressed in Refs. [9, 10, 13, 14], and Ref. [12] (for the present formalism). The key idea is to represent a quadratic function in ρ as a linear function on the tensor product $\rho\otimes\rho$:

$$o = \mathrm{Tr}\left(O\rho_{AB}\otimes\rho_{AB}\right). \quad \text{(D8)}$$

This function can be approximated by replacing $\rho\otimes\rho$ with a *symmetric* tensor product of two distinct snapshots $\hat{\rho}^{(i)}, \hat{\rho}^{(j)}$ $(i\neq j)$:

$$\tfrac{1}{2!}\sum_{\pi\in\mathcal{S}_2} \hat{\rho}_{AB}^{(\pi(i))}\otimes\hat{\rho}_{AB}^{(\pi(j))} = \tfrac{1}{2}\left(\hat{\rho}_{AB}^{(i)}\otimes\hat{\rho}_{AB}^{(j)} + \hat{\rho}_{AB}^{(j)}\otimes\hat{\rho}_{AB}^{(i)}\right). \quad \text{(D9)}$$

There are $\binom{M}{2}$ different ways of combining a collection of M snapshots $\hat{\rho}^{(1)},\dots,\hat{\rho}^{(M)}$ in this fashion. We can predict $o = \mathrm{Tr}(O\rho_{AB}\otimes\rho_{AB})$ by forming the empirical average over all of them:

$$\hat{o} = \binom{M}{2}^{-1}\sum_{i<j}\mathrm{Tr}\left(O\tfrac{1}{2!}\sum_{\pi\in\mathcal{S}_2}\hat{\rho}_{AB}^{(\pi(i))}\otimes\hat{\rho}_{AB}^{(\pi(j))}\right) = \binom{M}{2}^{-1}\sum_{i<j}\mathrm{Tr}\left(O_{(s)}\hat{\rho}_{AB}^{(i)}\otimes\hat{\rho}_{AB}^{(j)}\right). \quad \text{(D10)}$$

Here, we have implicitly defined the *symmetrization* $O_{(s)}$ of the original target function O. This ansatz is a special case of Hoeffding's U-statistics estimator [15]. Averaging boosts convergence to the desired expectation $\mathbb{E}\left[\hat{o}\right] = o$ and the speed of convergence is controlled by the variance (D6).

Restriction to subsystems is also possible. Suppose that O only acts nontrivially on a subsystem AB of both state copies. Then,

$$\hat{o} = \binom{M}{2}^{-1}\sum_{i<j}\mathrm{Tr}\left(O_{(s)}\hat{\rho}_{AB}^{(i)}\otimes\hat{\rho}_{AB}^{(j)}\right) \quad \text{(D11)}$$

and the effective problem dimension becomes $d_{AB}^2 = 4^{|AB|}$. The tensor product structure (D2) of the individual snapshots allows for generalizing linear variance bounds to this setting. Simply view $\hat{\rho}_{AB}^{(i)}\otimes\hat{\rho}_{AB}^{(j)}$ as a single snapshot of the quantum state $\rho_{AB}\otimes\rho_{AB}$. Fact 2 then ensures

$$\mathrm{Var}\left[\mathrm{Tr}\left(O_{(s)}\hat{\rho}_{AB}^{(i)}\otimes\hat{\rho}_{AB}^{(j)}\right)\right] \leq 4^{|AB|}\mathrm{Tr}\left(O_{(s)}^2\right). \quad \text{(D12)}$$

The full variance of $\hat{o}$ is controlled in part by this relation, but also features linear variance terms [12, App. 6.A]. Rather than reviewing this argument in full generality, let us focus on the task at hand: computing the variance associated with predicting the PT-moment of order two. Fix a bipartite subsystem AB of interest and rewrite p_2 as

$$p_2 = \mathrm{Tr}\left((\rho_{AB}^{(T_A)})^2\right) = \mathrm{Tr}\left(\rho_{AB}^2\right) = \mathrm{Tr}\left(\Pi_{AB}\rho_{AB}\otimes\rho_{AB}\right). \quad \text{(D13)}$$

Here, Π_{AB} denotes the swap operator that permutes the entire subsystems AB within two copies of the global system. We refer to Table I below for a visual derivation of this well-known relation. The swap operator is

symmetric under permuting tensor factors, Hermitian ($\Pi_{AB}^\dagger = \Pi_{AB}$) and orthogonal ($\Pi_{AB}^2 = \mathbb{I}_{AB}$). These properties ensure that the associated general estimator (D11) can be simplified considerably:

$$\begin{aligned}\hat{p}_2 =& \binom{M}{2}^{-1} \sum_{i<j} \mathrm{Tr}\left(\Pi_{AB} \hat{\rho}_{AB}^{(i)} \otimes \hat{\rho}_{AB}^{(j)}\right) \\ =& \binom{M}{2}^{-1} \sum_{i<j} \mathrm{Tr}\left(\hat{\rho}_{AB}^{(i)} \hat{\rho}_{AB}^{(j)}\right). \end{aligned} \tag{D14}$$

By construction, $\mathbb{E}[\hat{p}_2] = p_2 = \mathrm{Tr}(\rho^2)$ and the speed of convergence is controlled by the variance. This variance decomposes into a linear and a quadratic part. We expand the definition of the variance:

$$\begin{aligned}\mathrm{Var}[\hat{p}_2] &= \mathbb{E}\left[(\hat{p}_2 - \mathbb{E}[\hat{p}_2])^2\right] = \mathbb{E}\left[\hat{p}_2^2\right] - \mathbb{E}[\hat{p}_2]^2 \\ &= \binom{M}{2}^{-2} \sum_{i<j} \sum_{k<l} \left(\mathrm{Tr}\left(\hat{\rho}_{AB}^{(i)} \hat{\rho}_{AB}^{(j)}\right) \mathrm{Tr}\left(\hat{\rho}_{AB}^{(k)} \hat{\rho}_{AB}^{(l)}\right) - \mathrm{Tr}(\rho_{AB}^2)^2\right).\end{aligned} \tag{D15}$$

The size and nature of each contribution depends on the relation between the indices i, j, k, l [15]:

1. *all indices are distinct:* distinct indices label independent snapshots. In this case the expectation value factorizes completely and produces $\mathbb{E}\left[\mathrm{Tr}(\hat{\rho}_{AB}^{(i)} \hat{\rho}_{AB}^{(j)}) \mathrm{Tr}(\hat{\rho}_{AB}^{(k)} \hat{\rho}_{AB}^{(l)})\right] = \mathrm{Tr}(\rho_{AB}^2)^2$. This is completely offset by the subtraction of the expectation value squared. Hence, terms where all indices are distinct do not contribute to the variance.

2. *exactly two indices coincide:* In this case, the expectation value partly factorizes, e.g. $\mathbb{E}\left[\mathrm{Tr}(\hat{\rho}_{AB}^{(i)} \hat{\rho}_{AB}^{(j)}) \mathrm{Tr}(\hat{\rho}_{AB}^{(k)} \hat{\rho}_{AB}^{(j)})\right] = \mathbb{E}\left[\mathrm{Tr}(\rho_{AB} \hat{\rho}_{AB}^{(j)})^2\right]$ for $i \neq k$ and $j = l$. Such index combinations produce a linear variance term $\mathrm{Var}[\mathrm{Tr}(O\hat{\rho})]$ with $O = \rho_{AB}$. The entire sum contains $\binom{M}{2}\binom{2}{1}\binom{M-2}{2-1} = \binom{M}{2} 2(M-2)$ terms of this form.

3. *two pairs of indices coincide:* there are $\binom{M}{2}\binom{2}{2}\binom{M-2}{2-2} = \binom{M}{2}$ contributions of this form and each of them produces a quadratic variance $\mathrm{Var}[\mathrm{Tr}(O\hat{\rho}_{AB} \otimes \hat{\rho}'_{AB})]$ with $O = \Pi_{AB}$ (swap).

We conclude that the variance of $\hat{p}_2$ decomposes into linear and quadratic terms. These can be controlled via Rel. (D7) and Rel. (D12), respectively:

$$\begin{aligned}\mathrm{Var}[\hat{p}_2] =& \binom{M}{2}^{-1} \Big(2(M-2)\mathrm{Var}\left[\mathrm{Tr}(\rho_{AB}\hat{\rho}_{AB})\right] \\ &\qquad + \mathrm{Var}\left[\mathrm{Tr}(\Pi_{AB} \hat{\rho}_{AB}^{(1)} \otimes \hat{\rho}_{AB}^{(2)})\right]\Big) \\ \leq& \frac{4(M-2)2^{|AB|}}{M(M-1)} \mathrm{Tr}\left(\rho_{AB}^2\right) + \frac{2 \times 4^{|AB|}}{M(M-1)} \mathrm{Tr}\left(\Pi_{AB}^2\right) \\ \leq& 4\left(\frac{2^{|AB|} p_2}{M}\right) + 4\left(\frac{2^{1.5|AB|}}{M}\right)^2.\end{aligned} \tag{D16}$$

Chebyshev's inequality (D6) allows us to translate this insight into an error bound.

Lemma 2 (Error bound for estimating p_2)**.** *Fix a subsystem AB of interest and suppose that we wish to estimate $p_2 = \mathrm{Tr}\left((\rho_{AB}^{T_A})^2\right)$. For $\epsilon, \delta > 0$, a total of*

$$M \geq 8 \max\left\{\frac{2^{|AB|} p_2}{\epsilon^2 \delta}, \frac{2^{1.5|AB|}}{\epsilon\sqrt{\delta}}\right\} \tag{D17}$$

snapshots suffice to ensure that the estimator (D14) obeys $|\hat{p}_2 - p_2| \leq \epsilon$ with probability at least $1 - \delta$.

It is worthwhile to briefly discuss this two-pronged error bound. Asymptotically, i.e. for $M \to \infty$, the approximation error decays at a rate proportional to $1/\sqrt{M}$. This is the expected asymptotic decay rate for an estimation procedure that relies on empirical averaging (Monte Carlo). The actual rate is also multiplicative, i.e. the approximation error is proportional to the target p_2. In the practically more relevant, non-asymptotic setting, things can look strikingly different. For small and moderate sample sizes M, the variance bound (D16) is dominated by the next-to-leading order term ($2^{1.5|AB|} > 2^{|AB|} p_2$, especially if p_2 is small). Lemma 2 captures this discrepancy and heralds an error decay rate proportional to $1/M$ in this regime.

Finally, we point out that the dependence on δ in Eq. (D17) can be considerably improved by using median of means estimation [12]: split the total data into equally sized chunks, construct independent estimators and take their median. For this procedure, a sampling rate proportional to $\log(1/\delta)$ suffices. Moreover, median of means is much more robust towards outlier corruption and allows for using the same data to predict purities of many different subsystems simultaneously. This, however, comes at the price of somewhat larger constants in the error bound (D17) and heralds a tradeoff. In statistical terms, median of means estimation dramatically increases confidence levels ($1 - \delta$) at the cost of slightly larger error bars (confidence intervals). This tradeoff becomes advantageous when one attempts to predict very many properties from a single data set.

2. Predicting cubic properties (p_3 and $\mathrm{Tr}(\rho_{AB}^3)$)

Cubic properties can be predicted in much the same fashion as quadratic properties [12]. Write $o = \mathrm{Tr}\left(O\rho_{AB} \otimes \rho_{AB} \otimes \rho_{AB}\right)$ and approximate $\rho \otimes \rho \otimes \rho$ by a *symmetric* tensor product of three distinct snapshots $\hat{\rho}_{AB}^{(i)}, \hat{\rho}_{AB}^{(j)}, \hat{\rho}_{AB}^{(k)}$:

$$\tfrac{1}{3!} \sum_{\pi \in \mathcal{S}_3} \hat{\rho}_{AB}^{(\pi(i))} \otimes \hat{\rho}_{AB}^{(\pi(j))} \otimes \hat{\rho}_{AB}^{(\pi(k))}. \tag{D18}$$

There are $\binom{M}{3}$ different ways of combining a collection of M (independent) snapshots $\hat{\rho}_{AB}^{(1)}, \ldots, \hat{\rho}_{AB}^{(M)}$ in this fashion. We estimate the cubic function o by averaging over

all of them (U-statistics [15]):

$$\hat{o} = \binom{M}{3}^{-1} \sum_{i<j<k} \mathrm{Tr}\left(O \tfrac{1}{3!} \sum_{\pi\in\mathcal{S}_3} \hat{\rho}^{(\pi(i))} \otimes \hat{\rho}^{(\pi(j))} \otimes \hat{\rho}^{(\pi(k))} \right). \tag{D19}$$

Once more, the variance controls the rate of convergence to the desired target value $\mathbb{E}[\hat{o}] = \mathrm{Tr}(O\rho\otimes\rho\otimes\rho)$. This variance decomposes into a linear, a quadratic and a cubic part. The argument is a straightforward generalization of the analysis from the previous subsection. Rather than repeating the steps in full generality, we directly focus on the 3rd order PT-moment p_3 of a subsystem AB:

$$p_3 = \mathrm{Tr}\left((\rho_{AB}^{T_A})^3\right). \tag{D20}$$

For notational simplicity, we suppress the subscript AB indicating the subsystem of interest and label the shadows by lower-case indices: $\hat{\rho}_{AB}^{(i)} \mapsto \hat{\rho}_i$ for $i = 1,\dots,M$. Due to the cyclicity of the trace, the U-statistics estimator simplifies to

$$\begin{aligned}\binom{M}{3}\hat{p}_3 &= \sum_{i<j<k} \mathrm{Tr}\left(\tfrac{1}{3!} \sum_{\pi\in\mathcal{S}_3} \hat{\rho}_{\pi(i)}^{T_A}\hat{\rho}_{\pi(j)}^{T_A}\hat{\rho}_{\pi(k)}^{T_A} \right) \qquad \text{(D21)}\\ &= \sum_{i<j<k} \tfrac{1}{2}\left(\mathrm{Tr}\left(\hat{\rho}_i^{T_A}\hat{\rho}_j^{T_A}\hat{\rho}_k^{T_A}\right) + \mathrm{Tr}\left(\hat{\rho}_j^{T_A}\hat{\rho}_i^{T_A}\hat{\rho}_k^{T_A}\right)\right),\end{aligned}$$

where we have moved the normalization factor $\binom{M}{3}^{-1}$ to the left hand side in order to to increase readability. When computing the variance, we need to consider two sums over triples of distinct indices in $\{1,\dots,M\}$. If all indices are distinct, the overall contribution vanishes. Otherwise the contribution depends on the number $c \in \{1,2,3\}$ of indices the triples have in common. The number of distinct choices for two triples with exactly c integers in common is $\binom{M}{3}\binom{3}{c}\binom{M-3}{3-c}$ and we infer

$$\begin{aligned}&\binom{M}{3}\mathrm{Var}\left[\hat{p}_3\right]\\ =&\binom{3}{1}\binom{M-3}{2}\mathrm{Var}\left[\mathrm{Tr}\left((\rho^{T_A})^2\hat{\rho}_1^{T_A}\right)\right]\\ +&\binom{3}{2}\binom{M-3}{1}\mathrm{Var}\left[\mathrm{Tr}\left(\rho^{T_A}\tfrac{1}{2}\left(\hat{\rho}_1^{T_A}\rho_2^{T_A}+\hat{\rho}_2^{T_A}\hat{\rho}_1^{T_A}\right)\right)\right]\\ +&\mathrm{Var}\left[\tfrac{1}{2}\left(\mathrm{Tr}\left(\hat{\rho}_1^{T_A}\hat{\rho}_2^{T_A}\hat{\rho}_3^{T_A}\right)+\mathrm{Tr}\left(\hat{\rho}_2^{T_A}\hat{\rho}_1^{T_A}\hat{\rho}_3^{T_A}\right)\right)\right]\\ \le&\binom{M}{3}\left(\frac{9}{M}L+\frac{18}{M^2}Q+\frac{12}{M^3}C\right).\end{aligned} \tag{D22}$$

Here, $\hat{\rho}_1,\hat{\rho}_2,\hat{\rho}_3$ denote independent, random realizations of the snapshot $\hat{\rho}$ and we have introduced place-holders for linear (L), quadratic (Q) and cubic (C) contributions, respectively. For the task at hand, these contributions can be bounded individually and depend on the subsystem size AB:

1. *linear contribution:* set $O = (\rho_{AB}^{T_A})^2$ for notational brevity. We can use $\mathrm{Tr}(O\hat{\rho}^{T_A}) = \mathrm{Tr}(O^{T_A}\hat{\rho})$ to absorb the partial transpose in the linear function. Rel. (D7) then ensures

$$L \le 2^{|AB|}\mathrm{Tr}(\rho^2)^2, \tag{D23}$$

where we have also used $\mathrm{Tr}((O^{T_A})^2) = \mathrm{Tr}(O^2)$, as well as $\mathrm{Tr}(O^2) = \|O\|_2^2 \le \|O\|_1^2 = \mathrm{Tr}(O)^2$, because $O = \rho^2$ is psd.

2. *quadratic contribution:* We can bring $\frac{1}{2}\left(\mathrm{Tr}(\rho^{T_A}\hat{\rho}_1^{T_A}\hat{\rho}_2^{T_A}) + \mathrm{Tr}(\rho^{T_A}\hat{\rho}_2^{T_A}\hat{\rho}_1^{T_A}\right)$ into the canonical form $\mathrm{Tr}\left(O\hat{\rho}_{AB}^{(1)}\otimes\hat{\rho}_{AB}^{(2)}\right)$ by introducing

$$\begin{aligned}O = \tfrac{1}{2}&\left(\Pi_A(\rho\otimes\mathbb{I}_{AB})\Pi_B\right.\\ &\left.+\Pi_B(\rho\otimes\mathbb{I}_{AB})\Pi_A\right).\end{aligned} \tag{D24}$$

We refer to Table I for a visual derivation. Here, Π_A and Π_B are permutation operators that swap the two A- and B-systems, respectively. Rel. (D12) then ensures

$$Q \le 2^{2|AB|}\mathrm{Tr}\left(O^2\right) \le 2^{3|AB|}\mathrm{Tr}(\rho^2). \tag{D25}$$

The final estimate follows from exploiting $\Pi_A^2 = \Pi_B^2 = \mathbb{I}_{AB}$, as well as $\mathrm{Tr}\left(\rho^2\otimes\mathbb{I}_{AB}^2\right) = 2^{|AB|}\mathrm{Tr}(\rho^2)$.

3. *cubic contribution:* We can bring the cubic function $\frac{1}{2}\left(\mathrm{Tr}(\hat{\rho}_1^{T_A}\hat{\rho}_2^{T_A}\hat{\rho}_3^{T_A}) + \mathrm{Tr}(\hat{\rho}_2^{T_A}\hat{\rho}_1^{T_A}\hat{\rho}_3^{T_A})\right)$ into the canonical form $\mathrm{Tr}\left(O\hat{\rho}_1\otimes\hat{\rho}_2\otimes\hat{\rho}_3\right)$ by introducing

$$O = \tfrac{1}{2}\left(\overrightarrow{\Pi}_A\otimes\overleftarrow{\Pi}_B + \overrightarrow{\Pi}_A^\dagger\otimes\overleftarrow{\Pi}_B^\dagger\right), \tag{D26}$$

see Table I below. Here, $\overrightarrow{\Pi}_A$ is a cyclic permutation that exchanges all A-systems in a "forward" fashion ($A_1\mapsto A_2$, $A_2\mapsto A_3$, $A_3\mapsto A_1$), while $\overleftarrow{\Pi}_B$ is another cyclic permutation that exchanges all B-systems in a "backwards" fashion ($B_3\mapsto B_2$, $B_2\mapsto B_1$, $B_1\mapsto B_3$). A staightforward extension of Rel. (D12) to cubic functions implies

$$C \le 2^{3|AB|}\mathrm{Tr}(O^2) \le 2^{6|AB|}, \tag{D27}$$

because permutations are orthogonal ($\Pi\Pi^\dagger = \mathbb{I}$) and $\mathrm{Tr}(O^2)$ is dominated by $\mathrm{Tr}(\mathbb{I}_{AB}\otimes\mathbb{I}_{AB}\otimes\mathbb{I}_{AB}) = 2^{3|AB|}$.

Inserting these bounds into the variance formula for p_3 reveals

$$\begin{aligned}\mathrm{Var}\left[\hat{p}_3\right] \le& \frac{9}{M}L+\frac{18}{M^2}Q+\frac{12}{M^3}C \qquad \text{(D28)}\\ \le& 9\frac{2^{|AB|}}{M}\mathrm{Tr}(\rho^2)^2 + 18\frac{2^{3|AB|}}{M^2}\mathrm{Tr}(\rho^2) + 12\frac{2^{6|AB|}}{M^3}.\end{aligned}$$

Combining this insight with Chebyshev's inequality (D6) produces a suitable error bound. Recall that $p_2 = \mathrm{Tr}\left((\rho^{T_A})^2\right) = \mathrm{Tr}(\rho^2) \in [2^{-|AB|},1]$ denotes the purity of the subsystem in question.

Lemma 3 (Error bound for estimating p_3)**.** *Fix a subsystem AB of interest and suppose that we wish to estimate $p_3 = \mathrm{Tr}\big((\rho^{T_A})^3\big)$. For $\epsilon, \delta > 0$, a total of*

$$M \geq 39 \max\left\{ \frac{2^{|AB|} p_2^2}{\epsilon^2 \delta}, \frac{2^{1.5|AB|} p_2}{\epsilon\sqrt{\delta}}, \frac{2^{2|AB|}}{\epsilon^{2/3}\delta^{1/3}} \right\} \tag{D29}$$

snapshots suffice to ensure that the estimator (D21) obeys $|\hat{p}_3 - p_3| \leq \epsilon$ with probability at least $1 - \delta$.

This bound on the sampling rate provides different error decay rates for different regimes. For $M \to \infty$, the first term in the maximum dominates and the error decays at an asymptotically unavoidable rate proportional to $1/\sqrt{M}$. Conversely, for very small sample sizes M, the third term dominates and conveys a much larger decay rate proportional to $1/M^{3/2}$. In the intermediate regime, the second term may dominate and lead to a inverse linear decay rate $1/M$, instead. The dependence on the error parameter δ can once more be considerably improved (from $1/\delta$ to $\log(1/\delta)$) by using median of means estimation. This refinement also allows for using the same data to predict the cubic PT-moment of very many subsystems simultaneously [12].

Finally, we point out that the estimation error for $s_3 = \|\rho\|_3^3 = \mathrm{Tr}(\rho^3)$ can be bounded in exactly the same fashion. For $\epsilon, \delta > 0$, a sampling rate M that obeys Rel. (D29) also ensures that the U-statistics estimator U-statistics estimator

$$\hat{s}_3 = \binom{M}{3}^{-1} \sum_{i<j<k} \tfrac{1}{2} \left(\mathrm{Tr}\left(\hat{\rho}_i \hat{\rho}_j \hat{\rho}_k\right) + \mathrm{Tr}\left(\hat{\rho}_j \hat{\rho}_i \hat{\rho}_k\right) \right) \tag{D30}$$

obeys $|\hat{s}_3 - s_3| \leq \epsilon$ with probability $1 - \delta$.

The proof is almost identical to the p_3-analysis and we leave it as an exercise for the dedicated reader.

3. Additional numerical simulations

Here, we complement Fig. 2 of the MT by showing in Fig. D1 statistical errors in the estimation of p_2 and p_3 for the ground state of the transverse Ising model $H = J(\sum_i \sigma_i^x \sigma_{i+1}^x + \sigma_i^z)$ at criticality. We observe the same scaling behavior as in the case of the GHZ state. For p_2 [panel a)], there are indeed two regimes with different decay rates ($1/M$ and $1/\sqrt{M}$). For p_3 [panel b)], the latter two decay rates $1/M$ and $1/\sqrt{M}$ are also clearly visible. In contrast, the early regime decay rate is not as pronounced. This is likely due to limited system sizes – $1/M^{3/2}$ does appropriately capture the decay of red dots (largest system size considered) in the top left corner, but seems to be absent in decay rates for smaller system sizes.

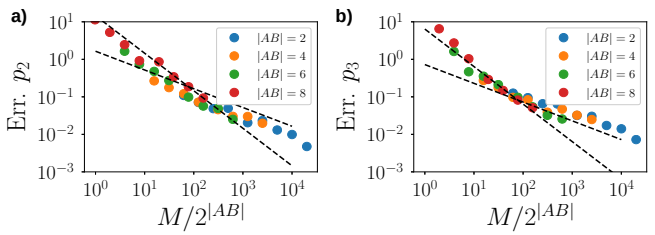


FIG. D1. Statistical errors for the ground state of the transverse field Ising model. Dashed lines represent scalings of $\propto 1/M$, and $\propto 1/\sqrt{M}$. In all cases, the number of measurements to estimate p_2 a) and p_3 b) with accuracy 0.1 is of the order of $100 \times 2^{|AB|}$.

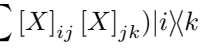
Appendix E: Auxiliary results and wiring diagrams

The arguments from the previous subsections make use of identities satisfied by traces of partial transposes of bipartite operators. Wiring diagrams – also known as tensor network diagrams – provide a useful pictorial calculus for deriving such identifies. We refer the interested reader to Refs. [16–18] for a thorough introduction and content ourselves here with a concise overview that will suffice for the purposes at hand. The wiring formalism represents operators as boxes with lines emanating from them. These lines represent contra- (on the left) and co-variant indices (on the right):

$$X = \sum_{i,j} [X_{ij}] \, |i\rangle\langle j| = \tag{E1}$$

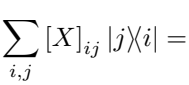
Two operators X and Y can be multiplied to produce another operator. This corresponds to an index contraction and is represented in the following fashion:

$$XY = \sum_{i,k} \Big(\sum_j [X]_{ij} \, [X]_{jk} \Big) |i\rangle\langle k| = \tag{E2}$$



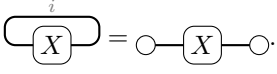
Transposition exchanges outgoing (contravariant) and incoming (covariant) indices

$$X^T = \sum_{i,j} [X]_{ij} \, |j\rangle\langle i| = \tag{E3}$$



while the trace pairs up both indices and sums over them:

$$\mathrm{Tr}(X) = \sum_i [X]_{ii} = \tag{E4}$$



We abbreviate this loop (contraction of leftmost and rightmost indices) by putting two circles at the end points of lines that should be contracted. This notation is not standard, but will considerably increase the readability of more complex contraction networks.

This basic formalism readily extends to tensor products if we arrange tensor product factors in parallel. For

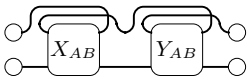
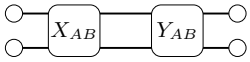
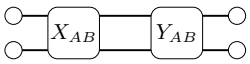
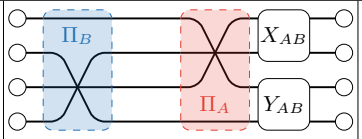
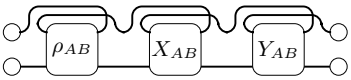
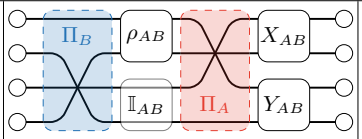
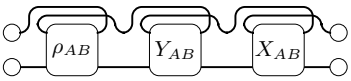
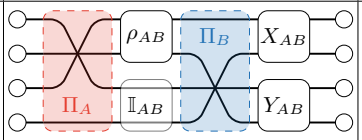
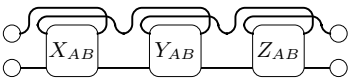
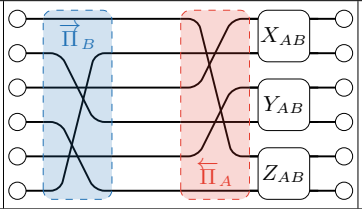
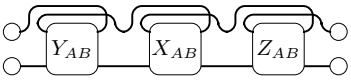
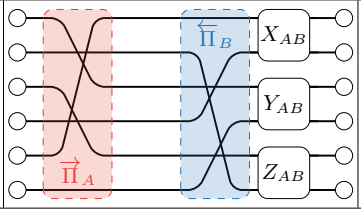
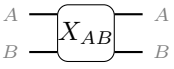
expression	diagram representation	diagram reformulation	modified expression
$\mathrm{Tr}(X_{AB}^{T_A} Y_{AB}^{T_A})$			$\mathrm{Tr}\left(X_{AB} Y_{AB}\right)$
$\mathrm{Tr}\left(X_{AB} Y_{AB}\right)$			$\mathrm{Tr}\left(\Pi_B \Pi_A X_{AB} \otimes Y_{AB}\right)$ Π_B, Π_A: swaps
$\mathrm{Tr}(\rho_{AB}^{T_A} X_{AB}^{T_A} Y_{AB}^{T_A})$			$\mathrm{Tr}\left(\Pi_B(\rho_{AB} \otimes \mathbb{I}_{AB})\Pi_A X_{AB} \otimes Y_{AB}\right)$
$\mathrm{Tr}(\rho_{AB}^{T_A} Y_{AB}^{T_A} X_{AB}^{T_A})$			$\mathrm{Tr}\left(\Pi_A(\rho_{AB} \otimes \mathbb{I}_{AB})\Pi_B X_{AB} \otimes Y_{AB}\right)$
$\mathrm{Tr}\left(X_{AB}^{T_A} Y_{AB}^{T_A} Z_{AB}^{T_A}\right)$			$\mathrm{Tr}\left(\overrightarrow{\Pi}_B \overleftarrow{\Pi}_A X_{AB} \otimes Y_{AB} \otimes Z_{AB}\right)$ $\overrightarrow{\Pi}_B, \overleftarrow{\Pi}_A$: cycle permutations
$\mathrm{Tr}\left(Y_{AB}^{T_A} X_{AB}^{T_A} Z_{AB}^{T_A}\right)$			$\mathrm{Tr}\left(\overleftarrow{\Pi}_A \overrightarrow{\Pi}_B X_{AB} \otimes Y_{AB} \otimes Z_{AB}\right)$ $\overleftarrow{\Pi}_A \overrightarrow{\Pi}_B = \left(\overrightarrow{\Pi}_B \overleftarrow{\Pi}_A\right)^\dagger$

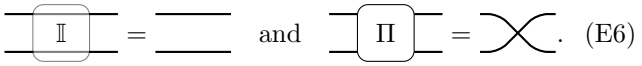
TABLE I. *Reformulations of relevant tensor product expressions:* The variance bounds in Sub. D 1 and Sub. D 2 are contingent on bringing certain expressions into canonical form, i.e. $\mathrm{Tr}\left(O X_{AB} \otimes Y_{AB}\right)$ for bilinear functions and $\mathrm{Tr}\left(O' X_{AB} \otimes Y_{AB} \otimes Z_{AB}\right)$ for trilinear ones. This table supports visual derivations for these reformulations. Expressions of interest (very left) are first translated into wiring diagrams (center left). Subsequently, the rules of wiring calculus are used to re-arrange the diagrams (center right). Translating them into formulas (very right) produces equivalent expressions that respect the desired structure.

instance, a bipartite operator features two parallel lines on the left and on the right:

$$X_{AB} = \quad (E5)$$



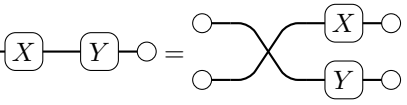
The upper lines represent the system A, while the lower lines represent system B. Two important bipartite operators are the identity $\mathbb{I}$ (do nothing) and the swap operator Π that exchanges the systems:



(E6)

Rules for multiplying and contracting operators readily extend to the tensor setting. This allows us to reformulate well-known expressions pictorially. For instance,

$$\mathrm{Tr}(XY) =$$

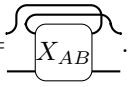


(E7)

$$= \mathrm{Tr}\left(\Pi X \otimes Y\right). \quad (E8)$$

The wiring formalism is also exceptionally well-suited to capture partial operations, like the partial transpose:

$$X_{AB}^{T_A} =$$



. (E9)

These elementary rules can be used to visually represent more complicated expressions – like a trace of multiple partial transposes. The wiring formalism provides a pictorial representation for such objects and a visual framework for modifying them. In particular, it is possible to bend, as well as unentangle, index lines and rearrange

tensor factors at will. Table I collects several such modifications that are important for the arguments above.

[1] K. Zyczkowski, Open Systems and Information Dynamics **10**, 297 (2003).
[2] S. Rana, Phys. Rev. A **87**, 054301 (2013).
[3] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, 2018).
[4] E. Wybo, M. Knap, and F. Pollmann, arXiv:2004.13072.
[5] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).
[6] R. Islam, R. Ma, P. M. Preiss, M. Eric Tai, A. Lukin, M. Rispoli, and M. Greiner, Nature **528**, 77 (2015).
[7] A. M. Kaufman, M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, P. M. Preiss, and M. Greiner, Science **353**, 794 (2016).
[8] N. M. Linke, S. Johri, C. Figgatt, K. A. Landsman, A. Y. Matsuura, and C. Monroe, Phys. Rev. A **98** (2018).
[9] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, Science **364**, 260 (2019).
[10] A. Elben, B. Vermersch, C. F. Roos, and P. Zoller, Phys. Rev. A **99**, 052323 (2019).
[11] M. Paini and A. Kalev, arXiv:1910.10543.
[12] H.-Y. Huang, R. Kueng, and J. Preskill, Nat. Phys. (2020).
[13] S. J. van Enk and C. W. J. Beenakker, Phys. Rev. Lett. **108**, 110503 (2012).
[14] A. Elben, B. Vermersch, M. Dalmonte, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **120**, 050406 (2018).
[15] W. Hoeffding, in *Breakthroughs in Statistics* (Springer, 1992) pp. 308–334.
[16] J. M. Landsberg, *Tensors: geometry and applications* (American Mathematical Society (AMS), 2012) p. 439.
[17] J. C. Bridgeman and C. T. Chubb, J. Phys. **A50**, 223001 (2017).
[18] R. Kueng, "Quantum and classical information processing with tensors (lecture notes)," (2019), Caltech course notes: `https://iqim.caltech.edu/classes`.