

# Quantum Tensor-Product Decomposition from Choi-State Tomography


Refik Mansuroglu<sup>1,\*</sup>, Arsalan Adil<sup>2</sup>, Michael J. Hartmann<sup>1</sup>, Zoë Holmes,<sup>3</sup> and Andrew T. Sornborger<sup>4</sup>

<sup>1</sup>Department of Physics, Friedrich-Alexander Universität Erlangen-Nürnberg (FAU), Staudtstraße 7, 91058 Erlangen, Germany

<sup>2</sup>Center for Quantum Mathematics and Physics, and Department of Physics and Astronomy, University of California, Davis, One Shields Avenue, Davis, California 95616, USA

<sup>3</sup>École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

<sup>4</sup>Information Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico 87544, USA

 (Received 14 February 2024; revised 24 April 2024; accepted 17 June 2024; published 10 July 2024)

The Schmidt decomposition is the go-to tool for measuring bipartite entanglement of pure quantum states. Similarly, it is possible to study the entangling features of a quantum operation using its operator-Schmidt or tensor-product decomposition. While quantum technological implementations of the former are thoroughly studied, entangling properties on the operator level are harder to extract in the quantum computational framework because of the exponential nature of sample complexity. Here, we present an algorithm for unbalanced partitions into a small subsystem and a large one (the environment) to compute the tensor-product decomposition of a unitary the effect of which on the small subsystem is captured in classical memory, while the effect on the environment is accessible as a quantum resource. This quantum algorithm may be used to make predictions about operator nonlocality and effective open quantum dynamics on a subsystem, as well as for finding low-rank approximations and low-depth compilations of quantum circuit unitaries. We demonstrate the method and its applications on a time-evolution unitary of an isotropic Heisenberg model in two dimensions.

DOI: [10.1103/PRXQuantum.5.030306](https://doi.org/10.1103/PRXQuantum.5.030306)

## I. INTRODUCTION

Entanglement is a defining feature of quantum theory [1]. The powerful capability of sharing information in a superposition of coupled states within a composite system still fascinates and puzzles physicists even 100 years after the advent of quantum physics. Entanglement is used as a fundamental resource for quantum computing and has launched an entirely new paradigm for information processing [2].

For a fixed Hilbert-space partition,  $\mathcal{H} \cong \mathcal{H}_A \otimes \mathcal{H}_B$ , the Schmidt decomposition of a pure state,  $|\psi\rangle$ , into tensor products,  $|\psi\rangle = \sum_{k=1}^r \sigma_k |a_k\rangle \otimes |b_k\rangle$ , reveals features about the shared entanglement between the two subsystems,  $\mathcal{H}_A$  and  $\mathcal{H}_B$ . A disentangled state, or tensor-product state, will consist of a single nonzero term, while an entangled state will have a Schmidt rank  $r > 1$ . Analogously, we

can define the tensor-product decomposition (TPD) [3] or operator-Schmidt decomposition [4] of a unitary operator,  $U$ , via

$$U = \sum_{k=1}^R s_k A_k \otimes B_k. \quad (1)$$

Here,  $A_k \in \mathcal{L}(\mathcal{H}_A)$ ,  $B_k \in \mathcal{L}(\mathcal{H}_B)$  are linear operators acting on the subsystems  $\mathcal{H}_{A/B}$  and the rank,  $R$ , is the minimal number of nonzero terms in the TPD. Without loss of generality, we can impose the  $A_k$  and  $B_k$  to be orthogonal with respect to the Hilbert-Schmidt inner product and normalized to  $\|A_k\|^2 = d_A := \dim(\mathcal{H}_A)$  and  $\|B_k\|^2 = d_B := \dim(\mathcal{H}_B)$  using the 2-norm  $\|\cdot\|$ . Note that  $A_k$  and  $B_k$  are not unitary, in general. Furthermore, the  $s_k$  are non-negative real numbers that are constrained to sum up to one, by unitarity of  $U$ , i.e.,  $\sum_k s_k^2 = 1$  (for details, see Appendix A).

As a theoretical tool, TPD has previously been used to classify the nonlocal and entangling content of unitaries [5–7] and also for analysis of the time evolution of quantum many-body systems [8] and quantification of quantum chaos [9,10]. Recently, TPD has been used to construct entanglement witnesses [11]. While these

\*Contact author: Refik.Mansuroglu@fau.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

advances motivate a systematic method to obtain the TPD of a quantum operator, current approaches are limited to classical resources.

Van Loan and Pitsianis [3,12] have developed a classical algorithm to find the TPD of an operator  $T$ , not necessarily unitary, using the singular-value decomposition of a reordered version of  $T$ . In quantum information processing, the operator of interest will typically require classical memory that grows exponentially in the number of qubits, making such classical methods inaccessible. While the measurement of Schmidt decompositions on quantum states has already been thoroughly studied [13–15], works on the operator level remain limited to specific problems that can be treated analytically [5,7,16].

Here, we bring the tensor-product decomposition into a quantum algorithmic framework. In particular, we present a hybrid quantum-classical algorithm that performs the quantum tensor-product decomposition (QTPD) described in Eq. (1) for a unitary matrix  $U$  with a known quantum circuit representation. If we assume an asymmetric split for which  $d_A \ll d_B$ , QTPD provides the operators  $A_k$  in classical memory, whereas the  $B_k$  are accessed as a quantum resource distilled out of  $U$ . The complete algorithm is visualized in Fig. 1.

We discuss a number of immediate applications and new directions for future research that are enabled with QTPD. Alongside low-rank approximations, QTPD provides a tool for studying entanglement, with application to entanglement witnesses and measures of entanglement generation [6,11], classically assisted simulation of (open) quantum dynamics [17–19], and low-depth compilation techniques [20]. The fact that the  $A_k$  are stored classically goes hand in hand with the philosophy of hybrid quantum computing, which is to use quantum resources as little as possible but in the most crucial step. We note, however, that QTPD collects all necessary data from the quantum computer at the start of the algorithm and does not require a hybrid quantum-classical optimization loop [21]. We demonstrate QTPD and its applications on the time-evolution operator of an isotropic Heisenberg model.

## II. QUANTUM TENSOR-PRODUCT DECOMPOSITION

We introduce QTPD in two steps. First, a matrix representation of the action on  $\mathcal{H}_A$  is captured via tomography and, second, we discuss a distillation technique

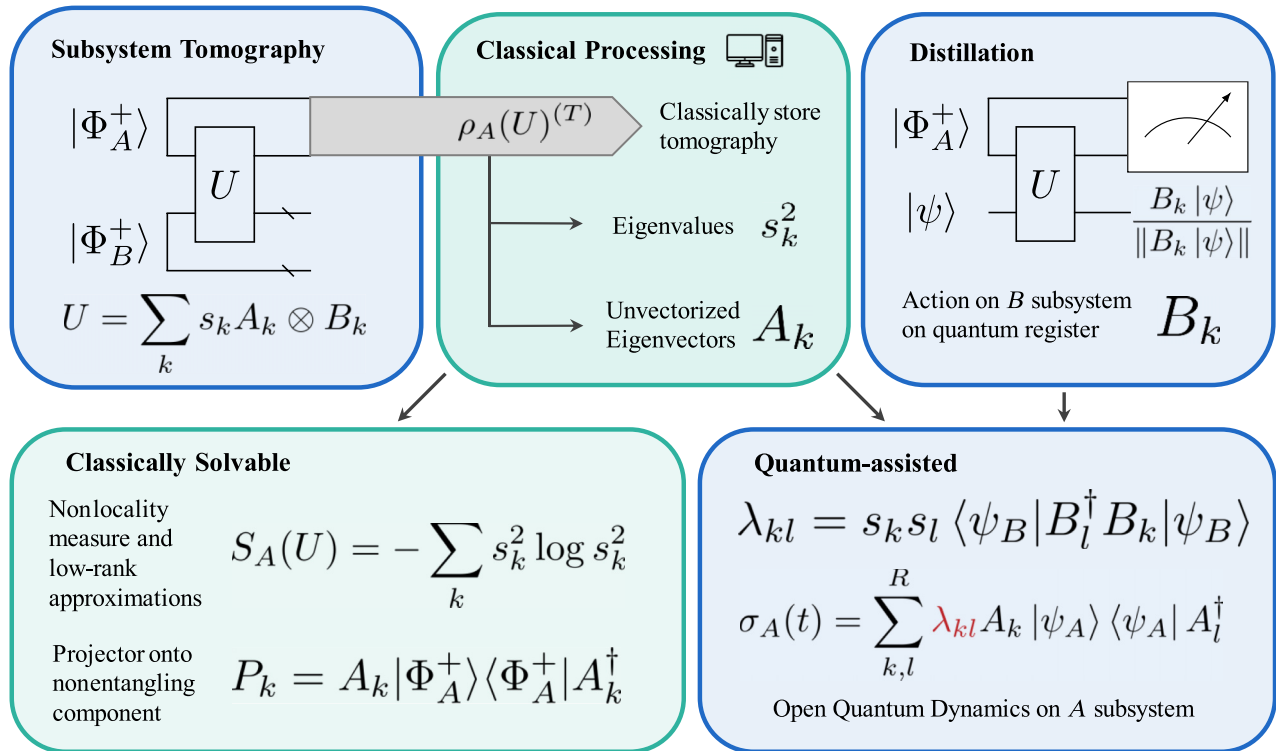


FIG. 1. A summary of quantum tensor-product decomposition (QTPD). Given a unitary operation  $U$  as a quantum resource, it is decomposed into the form of Eq. (1) in two steps. First, the Choi state of  $U$  is prepared and tomography is performed on the subsystem  $\mathcal{H}_A$ . The classical snapshot  $\rho_T$  of the state  $\rho$  of subsystem  $\mathcal{H}_A$  is then classically diagonalized to obtain the tensors  $A_k$  and the Schmidt values  $s_k$  [cf. Eq. (3)]. With this classical information, the nonlocality measure  $S_A(U)$  introduced in Eq. (17) can be calculated. The action of  $U$  on the environment via  $B_k$  is consequently obtained by a measurement of the observable  $P_k$  from Eq. (4) on the subsystem  $\mathcal{H}_A$ . The green boxes denote fully classical steps and the blue boxes denote steps where a quantum computer is used.

to subsequently capture the action on  $\mathcal{H}_B$  as a quantum channel. We finally compare the required resources for QTPD against its classical competitor, discuss possible adaptations for the near term, and comment on error propagation.

### A. The algorithm

We start with a unitary operator  $U$  that is accessible as a quantum resource (i.e., it is accessible as an oracle or its circuit representation is known). Our aim is to get a classical snapshot of the reduced action of  $U$  on  $\mathcal{H}_A$ , which is implicit in the operators  $A_k$  in Eq. (1). Consider the action of  $U$  on the two generalized Bell states

$$|\Phi_{A/B}^+\rangle = \frac{1}{\sqrt{d_{A/B}}} \sum_{i=1}^{d_{A/B}} |i\rangle|i\rangle, \quad (2)$$

which are states on two copies of the subsystems  $\mathcal{H}_A$  and  $\mathcal{H}_B$ , respectively (cf. the first circuit in Fig. 1). After tracing out  $\mathcal{H}_B$  from  $U(|\Phi_A^+\rangle \otimes |\Phi_B^+\rangle)$ , we are left with the mixed state of the vectorizations  $\text{vec}(A_k) = (1/\sqrt{d_A}) \sum_i^{d_A} |i\rangle A_k |i\rangle$ , i.e.,

$$\rho_A(U) = \frac{1}{d_A} \sum_k^R s_k^2 \text{vec}(A_k) \text{vec}(A_k)^\dagger, \quad (3)$$

which can be derived using the orthonormality of the  $A_k$  and  $B_k$  (see Appendix B 1). Unvectorizing the  $\text{vec}(A_k)$  is exponentially hard, in general [22]. Since  $d_A$  is assumed to be much smaller than  $d_B$ , a tomography of the state  $\rho_A(U)$  can be taken and stored as a classical snapshot. Diagonalizing  $\rho_A(U)$  finally yields the eigenvectors  $\text{vec}(A_k)$  with corresponding eigenvalues  $s_k^2$ . Note that this is mathematically equivalent to finding one half (on  $\mathcal{H}_A$ ) of the Schmidt decomposition of the Choi state of  $U$ .

The above algorithm not only yields information about  $s_k$  and  $A_k$  but we can also find  $B_k$  as a quantum resource. Once the  $A_k$  are found, one can distill out the individual  $B_k$ . That is, given a state  $|\psi\rangle$ , one can prepare  $B_k|\psi\rangle$ . This is done via a partial measurement of the projector

$$P_k = A_k |\Phi_A^+\rangle \langle \Phi_A^+| A_k^\dagger \quad (4)$$

on the state  $U(|\Phi_A^+\rangle \otimes |\psi\rangle)$  (cf. the second circuit in Fig. 1 and see Appendix B 2 for a derivation). The projectors,  $P_k$ , are orthogonal, i.e.,  $P_k P_l = \delta_{kl} P_k$ , which is a direct consequence of the orthonormality of the  $A_k$ . As a result, they can be simultaneously measured, such that every shot of the distillation circuit (cf. Fig. 1) yields the normalized output state  $B_k|\psi\rangle / \|B_k|\psi\rangle\|$  with probability  $p_k = s_k^2 \|B_k|\psi\rangle\|^2$ . Note that  $\sum_{k=1}^R p_k = 1$  from unitarity of  $U$  (for details, see Appendix B 2).

If one is interested in the action of a specific  $B_k$  only, the distillation process involves postselection on the outcome of the partial measurement. This comes with a sample overhead of  $\mathcal{O}(1/p_k) = \mathcal{O}(1/s_k^2 \|B_k|\psi\rangle\|^2)$ , which is never a serious issue. The overhead becomes large when either  $\|B_k|\psi\rangle\|$  or  $s_k$  become small. In the first case, the output state is close to the zero vector and in the second case, the sampled tensor component is a small contribution in a low-rank approximation of  $U$ .

QTPD can be used to determine such approximations to  $U$  of a specified rank. A set of operators  $\{C_k\}$  and  $\{D_k\}$ , such that the 2-norm

$$\left\| U - \sum_k^r t_k C_k \otimes D_k \right\| \quad (5)$$

is minimal, is called a rank- $r$  approximation. We have introduced the positive real-valued scalars  $t_k$  following the same convention as in the tensor-product decomposition of  $U$ . The special case  $r = 1$  from Eq. (5) corresponds to the well-known nearest Kronecker problem [3,12]. The solution to minimize Eq. (5) is the sum of product operators,  $\sum_k^r s_k A_k \otimes B_k$ , that correspond to the largest eigenvalues  $\{s_k^2\}_{k=1}^r$  of  $\rho_A(U)$  [cf. Eq. (3); for a proof, see Appendix C 1].

As the  $s_k$  and the  $A_k$  are classically stored but the  $B_k$  are not, we cannot classically store a low-rank approximation. A low-rank approximation can be used to suppress the sample complexity of QTPD whenever the sample budget is limited. This allows us to resolve just the singular values,  $s_k$ , that are sufficiently large and still provide a good approximation of  $U$ . In particular, if  $\varepsilon$  is the tolerable error of resolving the largest eigenvalues of  $\rho_A(U)$ , then the sample complexity scales as  $\mathcal{O}(R(d_A^2/\varepsilon^2))$  [23]. Hence, we achieve an  $\varepsilon$ -close approximation to  $\rho_A(U)$  in the operator norm by dropping every  $s_k^2 < \varepsilon$ . A low-rank approximation, in which the  $B_k$  are accessible on a quantum computer, as described above, can be seen as an application of QTPD. It differs from the applications that are discussed in Sec. III, as it is always applied along with tomography.

### B. Resources for QTPD

QTPD has the obvious advantage over classical methods that the unitary  $U$  can be loaded as a quantum circuit. Since the Bell-pair creation can be executed in depth 2, the depth of QTPD is primarily given by the depth of the circuit  $U$ . The run time of QTPD is hence dominated by the sample complexity of the tomography, which in this case is  $\tilde{\mathcal{O}}(R(d_A^2/\varepsilon^2))$  [23]. In memory, QTPD comes with a linear overhead of  $n$  ancilla qubits and requires classical memory to store  $\mathcal{O}(R d_A^2)$  complex numbers in the worst case.

The  $B$ -distillation step is similar in depth and admits a smaller sample complexity  $\mathcal{O}(s_{k_0}^{-2})$ , where  $s_{k_0}$  is the

smallest coefficient to be resolved. It also only needs  $n_A$  ancilla qubits and no additional classical memory.

A classical method to find the TPD of  $U$  goes back to van Loan and Pitsianis [3], who have shown that the operator  $\mathcal{R}(U) \in \text{Mat}(d_A^2 \times d_B^2, \mathbb{C})$  with permuted elements, such that

$$\mathcal{R}(U) = \sum_k s_k \text{vec}(A_k) \text{vec}(B_k)^\dagger, \quad (6)$$

encodes the tensor factors in its singular-value decomposition. The tensor factors  $A_k$  and  $B_k$  can be derived as the unvectorized left and right eigenvectors of  $\mathcal{R}(U)$  and the  $s_k$  are its singular values. If we neglect the run time for reshaping  $U$  into  $\mathcal{R}(U)$ , the bottleneck of the van Loan and Pitsianis algorithm comes from the numerical solution for the singular-value decomposition, which is  $\mathcal{O}(d_B^2 d_A^2 R)$  [24,25]. The factor  $R$  can be further weakened by randomization techniques [25].

This, together with the required classical memory to store the  $d^2$  complex elements of  $U$ , makes the classical algorithm uncompetitive for large dimensions  $d$ . To be precise, QTPD achieves a superpolynomial speed-up of a factor  $\tilde{\mathcal{O}}(d_B^2 \varepsilon^2)$  and memory savings by a factor of  $d_B^2/R$ , both of which grow exponentially in system size. If  $U$  is given as a sparse matrix, Krylov-subspace methods can be employed, which lower the required memory but not the run time [25].

### C. Circumvention of doubling the system size

Next to the inevitable scaling in subspace dimension [22], the greatest challenge within QTPD in the near term will be to keep the entangled Bell pairs coherent until measurement. Further, small quantum processors with fast readout and long coherence times are most efficient when memory requirements are traded off against run time.

To this end, the effect of the Bell pairs  $|\Phi_{A/B}^+\rangle$  is reduced to an average over a basis of  $\mathcal{H}_A$  and  $\mathcal{H}_B$ , respectively. In every run, a random initial state drawn from a basis of choice (e.g., the computational basis) is fixed. The output state for a fixed basis state  $|j_A\rangle$  of  $\mathcal{H}_A$  and  $|j_B\rangle$  of  $\mathcal{H}_B$  is

$$\sum_k s_k A_k |j_A\rangle \otimes B_k |j_B\rangle. \quad (7)$$

Averaging over the basis in  $\mathcal{H}_B$  accounts for an effective partial trace, using

$$\begin{aligned} \mathbb{E}_{|b\rangle}[\langle b|B_k B_l|b\rangle] &= \frac{1}{d_B} \sum_{j=1}^{d_B} \langle j|B_k B_l^\dagger|j\rangle \\ &= \frac{1}{d_B} \text{Tr}(B_k B_l^\dagger) = \delta_{kl}, \end{aligned} \quad (8)$$

where the expectation value is taken over the discrete set of basis states in  $\mathcal{H}_B$ . After tracing out  $\mathcal{H}_B$ , we are thus

left with

$$\sum_k s_k^2 A_k |j_A\rangle \langle j_A| A_k^\dagger. \quad (9)$$

If we keep the input-output relation within  $\mathcal{H}_A$ , the vectorization of the  $A_k$  can be reconstructed using tomography and summing over all basis states in  $\mathcal{H}_A$ :

$$\sum_{j=1}^{d_A} |j\rangle A_k |j\rangle = \sqrt{d_A} \text{vec}(A_k). \quad (10)$$

Recall the definition of the normalized vectorization above Eq. (3). As a result, the same state as in Eq. (3), coming from the parallelized version using the Choi state, can be reconstructed in classical postprocessing. As opposed to the Choi-state-based version, however, the run time of this sequential version of QTPD is increased by the repeated tomography for a fixed basis state  $|j\rangle \in \mathcal{H}_A$ , yielding an overhead factor of  $d_A$  in run time. Convergence to the mean value from averaging over  $\mathcal{H}_B$ , on the other hand, does not introduce an additional sample overhead, as it is equivalent to tracing out  $\mathcal{H}_B$  as part of the Choi-state-based approach.

### D. Error analysis

In general, the error from shot noise in tomography will be operator valued and can be viewed as the difference between the correct state,  $\rho_A(U)$ , and the classical snapshot,  $\rho_A(U)^{(T)}$ , i.e.,  $\varepsilon^{(T)} = \rho_A(U) - \rho_A(U)^{(T)}$ . The shot-noise error,  $\varepsilon^{(T)}$ , propagates through QTPD and thus introduces errors to the  $s_k$ ,  $A_k$ , and  $B_k$ . We discuss the technical details in Appendix D 1 and briefly present the results here. In order to understand how errors propagate, we consider the errors of the eigenvalues and eigenvectors of  $\rho_A(U)^{(T)}$  separately:

$$\begin{aligned} \rho_A(U) &= \rho_A(U)^{(T)} + \varepsilon^{(T)} \\ &= (V - \varepsilon^{(V)})(D - \varepsilon^{(D)})(V - \varepsilon^{(V)})^\dagger + \varepsilon^{(T)}. \end{aligned} \quad (11)$$

Our aim is to express the errors of  $s_k$  and  $A_k$  via  $\varepsilon = \max(\|\varepsilon^{(D)}\|, \|\varepsilon^{(V)}\|)$ . To this end, we define the error measures  $\varepsilon_k^{(S)} = s_k^2 - s_k^{(T)2}$  and  $\varepsilon_k^{(A)} = A_k - A_k^{(T)}$ . With these measures, we can trivially relate

$$\|\varepsilon^{(D)}\| = \|D - D^{(T)}\| = \sqrt{\sum_k (\varepsilon_k^{(S)})^2} \leq d_A \varepsilon^{(S)}, \quad (12)$$

where the factor  $d_A$  can be removed by appropriate normalization of the 2-norm in order to represent an average-case error. Less trivially, but after a straightforward calculation



(see Appendix D 1), we can also relate

$$\|\varepsilon_k^{(A)}\| = \|A_k - A_k^{(T)}\| = \sqrt{-2 \left| \langle A_k^{(T)} | \varepsilon_k^{(A)} \rangle \right|}, \quad (13)$$

and with this finally

$$\|\varepsilon^{(V)}\| = \|V - V^{(T)}\| = \sqrt{-2 \sum_{k,l} \left| \frac{\langle A_k^{(T)} | \varepsilon_l^{(A)} \rangle}{d_A} \right|}. \quad (14)$$

We can show that  $\|\varepsilon^{(T)}\| \leq 3\varepsilon$ , which completes the propagation from tomography error to the tensor factors  $A_k$ . Applying the faulty  $A_k^{(T)}$  for distillation of the effective action  $B_k^{(T)}$  as a quantum resource imposes further error propagation on the projector  $P_k$ , defined in Eq. (4). Taking into account appropriate (faulty) normalization factors, we can also bound the error of  $B_k^{(T)}$  in the following way:

$$\|(B_k - B_k^{(T)})|\psi\rangle\| \leq \left(1 + \frac{1}{\sqrt{2}}\right) \frac{1}{\sqrt{d_A}} \|\varepsilon_k^{(A)}\| + \mathcal{O}(\varepsilon^2). \quad (15)$$

Altogether, we show that the error contributions in  $s_k$ ,  $A_k$ , and  $B_k$  are linear in the 2-norm  $\|\varepsilon^{(T)}\|$ .

### III. APPLICATIONS

The quantum tensor-product decomposition allows us to find and store the tensor components  $A_k$  in classical resources and  $B_k$  in a quantum resource. With this, we can solve a number of tasks of interest.

#### A. Nonlocality

One application of QTPD lies in measuring the nonlocality of the action of  $U$ , also called operator entanglement entropy [8–10], which further bounds how much entanglement  $U$  generates. The vectorization of  $U$  allows for a mapping of operators to quantum states. On this space, we can employ entanglement entropy measures that are defined for states. Consider the vectorized operator

$$\text{vec}(U) = \sum_k^R s_k \text{vec}(A_k) \otimes \text{vec}(B_k). \quad (16)$$

If we trace out the  $B$  subsystem, we obtain exactly the mixed state of Eq. (3). The von Neumann entanglement entropy of this state reads

$$S_A(U) = - \sum_k^R s_k^2 \log s_k^2 \quad (17)$$

and is a measure of the nonlocality of the action of  $U$ . The nonlocality  $S_A(U)$ , sometimes referred to as the Schmidt

strength [7,16], can be determined classically after a successful QTPD and admits a linear contribution from the tomography error to leading order (cf. Appendix D 3).

Note that although the nonlocality of product operations vanishes,  $S_A(A \otimes B) = 0$ , Eq. (17) alone is not a good measure of entanglement generation. For instance, the SWAP gate, which maps product states to product states, reaches the maximal value for Eq. (17). To measure entanglement generation, one considers the entangling power of a circuit [6,26,27]. Several measures for entangling power have been proposed, two of which we discuss in Appendix E.

#### B. Mereology

If  $U$  is generated by a physical Hamiltonian, one might be interested in searching for a bipartite factorization (sometimes referred to as the tensor-product structure) of the global Hilbert space such that the two subsystems are decoupled, i.e.,  $U = U_A \otimes U_B$ . Concretely, consider a Hamiltonian describing two interacting subsystems,  $H = \sum_i H_i^A \otimes H_i^B$ , where  $H_i^A$  and  $H_i^B$  are operators acting on states describing subsystems  $A$  and  $B$  in a Hilbert space factorized as  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ . Since this factorization is essentially a particular choice of a global basis, it can be related to another one by a unitary [28],  $\mathcal{H}_A \otimes \mathcal{H}_B \xrightarrow{V} \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$ , where  $V$  is a (nonlocal) unitary and states in  $A'$  and  $B'$  describe physically different subsystems than  $A$  and  $B$ . In particular, it may be possible to find a factorization of  $\mathcal{H}$  such that the Hamiltonian is decoupled, i.e.,  $VHV^\dagger = H_{A'} + H_{B'}$ .

Some have used this approach, with the goal of minimizing the interaction Hamiltonian between two subsystems, to understand the emergence of classicality [29]. Practically, this approach appears in cases in which taking a certain transformation can lead to analytically tractable forms of the Hamiltonian, e.g., in the case of the Jordan-Wigner transformation that transforms certain interacting qubit Hamiltonians to a set of free fermionic operators. While QTPD does not itself find the optimal basis  $V$  that will lead to approximately decoupled dynamics, it can be used to evaluate the cost function as part of another algorithm (such as the one proposed in Ref. [30]). Two candidates to minimize are  $\sum_{k=2}^R s_k$  or  $1 - s_1^2$ , where the  $s_k$  are the singular values in the tensor-product decomposition of the time propagator  $U = Ve^{-iH}V^\dagger = \sum_k^R s_k A_k \otimes B_k$ . The existence of such a *decoherence-free* split [31–33] is tightly connected to the spectral properties of  $U$  (see Appendix F 1 for an example and Appendix F 2 for a necessary and sufficient condition).

#### C. Fast quantum transform and classical simulability

A mereology algorithm can be further utilized to find an efficient compilation of a target unitary  $U$ . If there exists a basis in which  $U$  decouples,  $U = V^\dagger(U_A \otimes U_B)V$  can be implemented with a single layer after rotating into the basis

$V$ . A divide-and-conquer approach successively reduces the action of  $U$  into a tensor product of  $M$  local gates, i.e.,  $VUV^\dagger = U_{A^{(1)}} \otimes U_{A^{(2)}} \otimes \cdots \otimes U_{A^{(M)}}$ . Such a fast quantum transform requires a rotation into the basis  $V$ , which is entangling, in general.

More generally, for an arbitrary  $U$ , the closest fast quantum transform can be found via iterative QTPD, which can be performed efficiently if there is a single dominant coefficient  $s_{1\dots 1}$  in the multipartite factorization

$$U = \sum_{j_1, \dots, j_M}^{R_1, \dots, R_M} s_{j_1 \dots j_M} A_{j_1}^{(1)} \otimes \cdots \otimes A_{j_M}^{(M)}. \quad (18)$$

We have used the letter  $A$  for all operators to emphasize that the local dimensions are small enough to be classically simulated.

We can use the nearest unitary representations  $U_m$  of the tensor factors  $A_1^{(m)}$  of the rank-1 approximation of  $U$  to construct a fast quantum transform approximation to  $U$ . We show that the error is  $\mathcal{O}(\sqrt{1 - s_{1\dots 1}})$ . For this, we separate the error of the fast quantum transform into two terms, one representing the error from truncating all terms from  $U$  except  $s_{1\dots 1}$  and the other capturing the error by nearest unitary approximation  $(1/\sqrt{2d_{A^{(m)}}}) \|A_1^{(m)} - U_m\| = \varepsilon^{(m)} > 0$ . Writing  $1 - s_{1\dots 1} = \varepsilon_s > 0$ , we obtain

$$\begin{aligned} \frac{1}{\sqrt{2d}} \|U - U_1 \otimes \cdots \otimes U_M\| &\leq \sqrt{\varepsilon_s} \\ &+ \sqrt{\frac{1}{2}\varepsilon_s^2 + \sum_{m=1}^M (\varepsilon^{(m)})^2} \end{aligned} \quad (19)$$

(for details, see Appendix C 2).

If there is not only one but a polynomial number of dominant coefficients, a better approximation to the action of  $U$  can be achieved through a rank- $r$  approximation. In this case, the probability of sampling from the dominant  $B_k$  is suppressed by a polynomial factor. The simulation of a fast quantum transform,  $U_{A^{(1)}} \otimes U_{A^{(2)}} \otimes \cdots \otimes U_{A^{(M)}}$ , is not only classically efficient but it is also made up of low-entangling transformations. Instead of achieving a close unitary approximation, the goal here is to bound the bond dimensions necessary for a faithful tensor-network representation.

Since the nonlocality measure  $S_A(U)$  bounds the entanglement generation, it can be used as a witness to scan for clusterings of the Hilbert space with low entanglement between the clusters. If a cluster  $\mathcal{H}_{A^{(1)}}$ , on which the action of  $U$  is low entangling, is found, the entanglement generation between a second cluster  $\mathcal{H}_{A^{(2)}}$  and its environment  $\mathcal{H}_{A^{(3)}} \otimes \cdots \otimes \mathcal{H}_{A^{(M)}}$  will be bounded (cf. Fig. 2) as well. If the nonlocality of the full unitary  $U$  is bounded, it can thus be written as a matrix-product operator [34], of which the

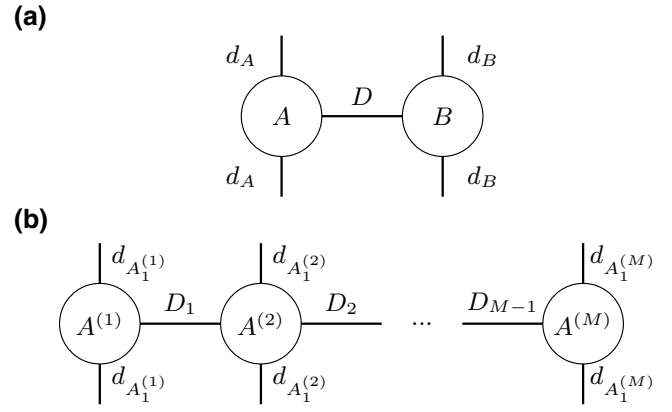


FIG. 2. Low-entanglement clustering for a matrix-product-state representation. (a) Upon division into subsystems  $A$  and  $B$ , a matrix-product operator representing  $U$  requires a certain bond dimension  $D \leq d_A^2$ , which is upper bounded by the lower dimension  $d_A^2 \leq d_B^2$  and dependent on the entanglement between  $A$  and  $B$ . (b) The multipartite decomposition of  $U$  into  $A_{j_m}^{(m)}$  allows for estimating the necessary bond dimensions  $D_m$ , which can be deduced by a low-rank approximation of the multipartite decomposition [cf. Eq. (18)].

fast quantum transform is an extremal case. This allows for an efficient classical representation of the output of  $U$ , e.g., via matrix-product states [35] or projected entangled-pair states [36].

Fast quantum transforms have conceptual similarities to entanglement forging [37], which is used to simulate a larger system by simulating the subsystems separately on a smaller quantum chip, if there are only few connecting gates in the compilation of  $U$ . As opposed to QTPD, these methods are typically concerned with symmetric splittings and aim for a reduction of quantum resources in the simulation of  $U$  instead of finding classically simulable subsystems.

#### D. Open quantum dynamics

QTPD is applicable to studying entangling dynamics or the decoherence of subsystem  $A$  into subsystem  $B$ . If we start with a product state  $|\psi_A\rangle \otimes |\psi_B\rangle$ , the evolved state within subsystem  $A$  will be mixed. The effective open quantum dynamics can be written in the form

$$\begin{aligned} \sigma_A &= \text{Tr}_B (U|\psi_A\rangle \otimes |\psi_B\rangle \langle \psi_A| \otimes \langle \psi_B| U^\dagger) \\ &= \sum_{k,l}^R \lambda_{kl} A_k |\psi_A\rangle \langle \psi_A| A_l^\dagger, \end{aligned} \quad (20)$$

where  $\lambda_{kl} = s_k s_l \langle \psi_B | B_l^\dagger B_k | \psi_B \rangle$ . While the operators  $A_k$  and the state  $|\psi_A\rangle$  can be stored on a classical machine, the  $\lambda_{kl}$  are not *a priori* accessible. Instead, the overlaps  $\langle \psi_B | B_l^\dagger B_k | \psi_B \rangle$  have to be determined using modified

Hadamard tests or SWAP tests [4,38] with different outputs of the  $B_k$  distillation via projective measurement of Eq. (4).

Once the  $\lambda_{kl}$  and  $A_k$  are stored classically, it is possible to simulate the open dynamics via Eq. (20) for any initial state  $|\psi_A\rangle$ . Equation (20) can be transformed into its Kraus representation, e.g., by diagonalizing the Choi matrix. In this manner, QTPD can be used as a quantum enhanced classical simulation algorithm [21] for open-system simulation. That is, quantum hardware is crucial to obtaining the  $A_k$  but then Eq. (20) acts as a classical surrogate to simulate the dynamics of any initial state and observable. Note the parallels with process learning [39,40], which aims to learn a quantum channel from measurements of local observables using classical resources, such as neural networks.

The error in predicting observables on  $\sigma_A$  can be bounded by the trace norm to the faulty prediction  $\sigma_A^{(T)}$  from tomography, which scales linearly with the tomography error  $\|\varepsilon^{(T)}\|$  as we show in Appendix D3. A naive quantum simulation with fixed initial state and fixed observable suffers from shot noise that has the same scaling in samples as  $\varepsilon^{(T)}$ .

#### IV. NUMERICAL EXPERIMENT

We demonstrate QTPD on a Hamiltonian-simulation problem for the isotropic Heisenberg model. To this end, we numerically solve the tensor decomposition by exact diagonalization of the unitary time evolution generated by the Hamiltonian

$$H = -J \sum_{\langle i,j \rangle} (X_i X_j + Y_i Y_j + Z_i Z_j), \quad (21)$$

with the Pauli matrices  $\{X, Y, Z\}$  and the sum over nearest neighbors denoted by  $\langle i,j \rangle$ . We discuss one system of toy size, the dynamics of which we can analytically solve, and a separate system on a two-dimensional (2D) grid that is small enough to be checked by exact diagonalization.

Let us consider a model of two qubits first. We show in Appendix H that the time-evolution operator generated by the Heisenberg Hamiltonian incorporates an oscillation between the identity and the SWAP operator for certain times. In between those times, entanglement is alternately generated and reduced. It is thus natural to view each qubit as a subsystem and study the open dynamics on one of the two qubits. Larger systems can be split in two and the swapping of excitations from  $\mathcal{H}_A$  to  $\mathcal{H}_B$  and vice versa can be studied, as well.

When more than two qubits are exchanging excitations, the overall dynamics are an ensemble of interfering oscillations, which depend on the initial state and the geometry of the interaction graph. To study both the distribution of excitations and the entanglement between the split into subsystems  $\mathcal{H}_A$  and  $\mathcal{H}_B$ , we stroboscopically measure the total magnetization  $M(\sigma_A(t))$  of subsystem  $\mathcal{H}_A$ , as well as

the von Neumann entanglement entropy of the output state  $S(\sigma_A(t))$ ,

$$M_A(\sigma_A(t)) = \text{Tr} \left( \sum_{i \in \mathcal{I}_A} Z_i \sigma_A(t) \right) \quad (22)$$

$$S(\sigma_A(t)) = -\text{Tr}(\sigma_A(t) \log \sigma_A(t)), \quad (23)$$

where  $\mathcal{I}_A$  denotes the index set for qubits in subsystem  $\mathcal{H}_A$ . For the sake of a clear comparison, we normalize all observables to take values between 0 and 1. This means that we divide the state entanglement entropies by  $\log(d)$  and the operator nonlocalities by  $\log(d^2)$ , where  $d$  is the dimension of the Hilbert space. The magnetization is transformed into an occupation number  $M_A \mapsto \frac{1}{2} - (M_A/2n_A)$ , where  $n_A$  denotes the number of qubits in subsystem  $\mathcal{H}_A$ .

In Fig. 3(a), we show the results for the two-qubit system. The nonlocality reaches its maximal value twice during the time interval  $t \in [0, (\pi/J)]$ , at which point the time-evolution operator oscillates between the SWAP and the identity operator. In between,  $U$  generates entanglement on the trial state, which is shown in green. We use QTPD to classically simulate the open quantum dynamics following Eq. (20). Starting with the initial state  $|1\rangle \otimes |0\rangle$ , we simulate the time evolution of the density matrix describing qubit 1 and measure the magnetization  $M_A$  and the entanglement entropy  $S$  from Eqs. (22) and (23). The excitation transfer between the two qubits is reflected in the oscillation of the magnetization between the extremal values 0 and 1. At these points, the entanglement entropy reaches zero, indicating an oscillatory SWAP between  $|10\rangle$  and  $|01\rangle$ . The data from QTPD are in exact agreement with the analytical expressions derived in Appendix H.

On a  $3 \times 2$  qubit grid on which excitations can swap between neighboring qubits, the overall dynamics are more complicated. While the nonlocality of the time-evolution operator quickly rises close to the maximum value, it no longer returns to zero within the time interval  $t \in [0, (4\pi/J)]$ . The trial initial state,  $|110000\rangle$ , that is evolved in time, does not return to a product state in the considered time interval, which shows that part of the nonzero nonlocality is in fact entanglement generated by  $U$ . We discuss two measures of entangling power (see Appendix E) on the example of the Heisenberg model in Appendix H, in order to specify the relation between nonlocality and the entangling properties of  $U$ .

#### V. CONCLUSIONS

In typical problems of quantum information processing, we are given a quantum circuit unitary  $U$  as a quantum resource. A quantum tensor-product decomposition enables the separation of a small subsystem  $\mathcal{H}_A$  from its environment and captures the effective action

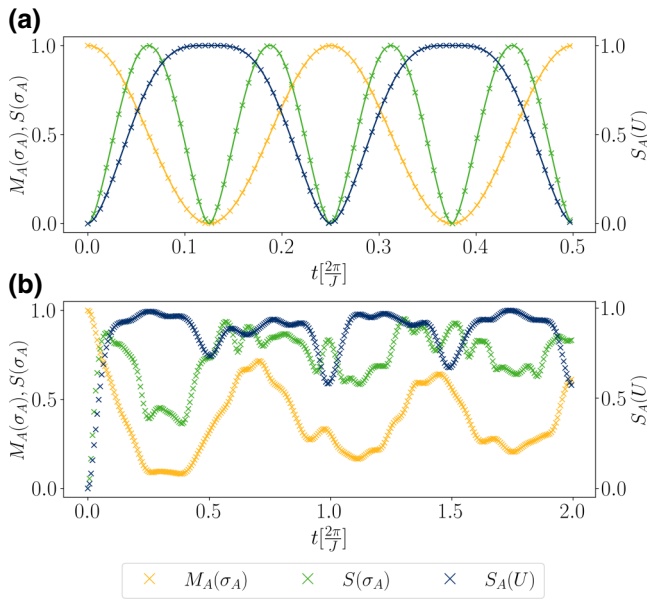


FIG. 3. The open quantum dynamics for the isotropic Heisenberg model. We study a subsystem of (a) one of two neighboring qubits and (b) two qubits of a two-dimensional (2D) grid of  $3 \times 2$  qubits. The blue points show the nonlocality measure  $S_A(U)$  for the respective time-evolution operator, while yellow and green contain the total magnetization  $M_A(\sigma_A)$  and the entanglement entropy  $S(\sigma_A)$  of the time-evolved state. The initial state of the total system reads  $|1\rangle \otimes |0\rangle$  in (a) and  $|11\rangle \otimes |0000\rangle$  in (b). In (a), the solid curves represent analytical predictions derived in Appendix H.

of  $U$  on  $\mathcal{H}_A$ . This allows one to classically predict the entanglement features of  $U$  and postprocess mereology and short-depth-compilation algorithms. Furthermore, it enables the study of open quantum dynamics interpreting the two subsystems,  $\mathcal{H}_A$  and  $\mathcal{H}_B$ , as system and environment, respectively. A generalization to decompose arbitrary tensors does not seem straightforward (see Appendix G). Using block-encoding techniques, QTPD could be applied to arbitrary tensors. A hurdle that arises in that case is to tame the sample complexity when sampling from the ancillary qubits needed for the block encoding of matrix-product operators [41]. We remain curious about extending QTPD to include quantum states (possibly via block encoding), where the sum of the QTPD coefficients,  $\sum_k s_k$ , may be used as a criterion to detect bipartite entanglement (see “realignment criterion” in, e.g., Refs. [1,42,43]) or verify matrix-product-operator structures of density operators [44].

QTPD paves the way for entanglement investigations at the operator level. A natural next step is to combine QTPD with an iterative search for quasiclassicality emerging in quantum systems. The nonlocality that upper bounds the entangling power can be used as a cost function to minimize the growth of entanglement with the

environment. Also, in reverse, QTPD can be used to verify decoherence-free structures [31].

Depending on the available resources in memory and connectivity, either the preparation of Bell pairs in depth 2 or the sequential approach can be used in a near-term application including QTPD. For medium-term quantum computing, a doubling in qubit number is within the scaling plans of modern quantum computing architectures, which typically strive for exponential growth. We can thus foresee an application in the near future that, e.g., utilizes 120 qubits (or 60 qubits in sequence) to solve the open dynamics of a  $(15 + 45)$ -qubit system, a simulation problem that is no longer accessible with classical methods. One such application is the integration of QTPD into dynamical mean-field methods, e.g., for the simulation of impurity models. We leave these directions for future research.

Alternative process-learning methods [39,40,45] make use of quantum machine-learning-assisted postprocessing. While a comparison of the efficiency is not immediate, QTPD provides a stable solution using tomography instead of quantum optimization.

Similar to process-learning approaches, QTPD can be generalized to capture the action of an arbitrary quantum channel  $\mathcal{E}$  using its Choi state. Quantum circuits on near-term hardware will inevitably suffer from noise and therefore a QTPD on a noisy circuit is a natural next step. If the hardware allows for the *intended* simulation of a specified quantum channel  $\mathcal{E}$ , QTPD along with its applications can be straightforwardly generalized, substituting the unitary  $U$  by  $\mathcal{E}$ .

## ACKNOWLEDGMENTS

R.M. thanks Norbert Schuch, Martin Larocca, and Dhruvil Patel for valuable discussions. This work received support from the German Federal Ministry of Education and Research via the funding program “Quantum Technologies—From Basic Research to the Market,” under Contract No. 13N16067 “Efficient Quantum Algorithms for the Hubbard Model”. It is also part of the Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus. Z.H. acknowledges support from the Sandoz Family Foundation—Monique de Meuron program for Academic Promotion. A.A. and A.T.S. acknowledge support from the U.S. Department of Energy (DOE), Office of Science, Office of High Energy Physics, QuantISED program. This work was supported by the U.S. Department of Energy (DOE) through a quantum computing program sponsored by the Los Alamos National Laboratory Information Science and Technology Institute. The LANL approval code for this paper is LA-UR-24-21093.



## APPENDIX A: AMBIGUITIES IN THE TENSOR DECOMPOSITION

The tensor-product decomposition is nonunique. Every tensor-product basis representation of operators in  $\mathcal{H}_{A/B}$  gives a decomposition of the form of Eq. (1), not necessarily, but potentially, with minimal rank,  $R_{\min}$ . Using the Gram-Schmidt theorem, we can assume the basis  $\{A_k \otimes B_k\}_k$  that appears in Eq. (1) to be orthogonal with respect to the Hilbert-Schmidt inner product, i.e.,

$$\begin{aligned} \langle A_k \otimes B_k | A_l \otimes B_l \rangle &= \langle A_k | A_l \rangle \langle B_k | B_l \rangle \\ &\stackrel{!}{=} \delta_{kl} \|A_k\|^2 \|B_k\|^2. \end{aligned} \quad (\text{A1})$$

Note that this ensures orthogonality of some of the  $A_k$  or  $B_k$  but not all. Let us assume for now that  $\langle B_k | B_l \rangle = \delta_{kl} \|B_k\|^2$ , so that Eq. (A1) is satisfied. A second symmetry is scale invariance,  $A_k \otimes B_k = (\lambda A_k) \otimes ((1/\lambda) B_k)$ , for  $\lambda \in \mathbb{C}$ . With this, we can, e.g., fix the norms of the  $B_k$  to be equal to the dimension of  $\mathcal{H}_B$ , i.e.,  $\|B_k\|^2 = d_B$ . Moreover, we can rotate the  $A_k$  and simultaneously counter-rotate the  $B_k$  with a linear superoperator  $L : \text{span}\{A_k\} \rightarrow \text{span}\{A_k\}$  that maps  $L(A_k) = \sum_j L_{jk} A_j$ .  $L$  can also be interpreted as a superoperator on  $\text{span}\{B_k\}$  with the analogue action  $L(B_k) = \sum_j L_{jk} B_j$ . If we assume  $L$  to be unitary, i.e.,  $\sum_k L_{ik} L_{jk}^* = \sum_k L_{ki} L_{kj}^* = \delta_{ij}$ , we can straightforwardly see that

$$\begin{aligned} U &= \sum_k A_k \otimes B_k = \sum_k L^\dagger(L(A_k)) \otimes B_k \\ &= \sum_{i,j,k} L_{ji}^* L_{jk} A_i \otimes B_k \\ &=: \sum_j \tilde{A}_j \otimes \tilde{B}_j \end{aligned} \quad (\text{A2})$$

is also a tensor decomposition. We have defined  $\tilde{A}_j := \sum_i L_{ji}^* A_i$  and  $\tilde{B}_j := \sum_i L_{ji} B_i$ . Whenever the boundary of the sum is omitted, we understand the sum to go through  $k \in \{1, \dots, R\}$ . Unitarity of  $L$  implies that the orthogonality of the  $B_k$  remains. Further, one can choose  $L$  such that the  $A_k$  are also orthogonal. To see this, define the linear operator  $\mathcal{O}_{kl} = \langle A_k | A_l \rangle$  and the transformed version  $\tilde{\mathcal{O}}_{kl} = \langle \tilde{A}_k | \tilde{A}_l \rangle$ . Inserting the definition of  $\tilde{A}_k$ , we obtain

$$\tilde{\mathcal{O}}_{kl} = \sum_{i,j} L_{ki} L_{lj}^* \mathcal{O}_{ij} = (L \mathcal{O} L^\dagger)_{kl}. \quad (\text{A3})$$

Since  $\mathcal{O}$  is Hermitian, we can choose  $L^\dagger$  to consist of the eigenvectors of  $\mathcal{O}$ , such that with this choice

$$\tilde{\mathcal{O}}_{kl} = \langle \tilde{A}_k | \tilde{A}_l \rangle = \delta_{kl} \|A_k\|^2. \quad (\text{A4})$$

With slight abuse of notation, we omit the tilde in the following and store the information about the norms  $\|A_k\|$  separately in scalars  $s_k = (\|A_k\|/d_A)$ , which normalizes the  $A_k$  accordingly. Finally, we can also choose the  $s_k$  to be non-negative real numbers, as any complex phase can be absorbed into the  $B_k$ , e.g.,

$$s_k e^{i\phi_k} A_k \otimes B_k = s_k A_k \otimes (e^{i\phi_k} B_k). \quad (\text{A5})$$

This transformation does not alter the norms and orthogonality of the  $B_k$ . We have now chosen a specific tensor decomposition  $U = \sum_k s_k A_k \otimes B_k$ , for which we can assume the following without loss of generality:

- (1) The operators  $A_k$  and  $B_k$  are orthogonal with respect to the Hilbert-Schmidt inner product.
- (2) The  $A_k$  and  $B_k$  are normalized, such that  $\|A_k\|^2 = d_A$  and  $\|B_k\|^2 = d_B$ .
- (3) The  $s_k$  are non-negative real numbers.

By unitarity of  $U$ , the  $s_k$  are further constrained to sum up to one, i.e.,  $\sum_k s_k^2 = 1$ .

## APPENDIX B: TWO-STEP QUANTUM TENSOR-PRODUCT DECOMPOSITION

In the following, we provide technical details for the proposed algorithm. QTPD involves two steps. First, the operators  $A_k$  acting on the smaller subsystem are obtained classically together with the coefficients  $s_k$  via diagonalization of a tomographic snapshot of a Choi state. In the second step, this information is used to construct a projective measurement that allows the distillation of  $B_k$  out of  $U$ .

### 1. Classical snapshot state

The first step of QTPD involves a tomography of the density matrix  $\rho_A(U)$  of the Choi state of  $U$  reduced to subsystem  $A$ , which we calculate in the following. First, we define the full Choi state:

$$\begin{aligned} |\Phi(U)\rangle &:= (\mathbb{I} \otimes U \otimes \mathbb{I}) |\Phi_A^+\rangle |\Phi_B^+\rangle = \sum_k \sum_{i_A, i_B} s_k |i_A\rangle \\ &\otimes A_k |i_A\rangle \otimes |i_B\rangle \otimes B_k |i_B\rangle. \end{aligned} \quad (\text{B1})$$

After tracing out  $\mathcal{H}_B$ , we obtain



state is thus the following mixed state:

$$\mathcal{E}[\rho] = \sum_m s_m^2 (A_m |\Psi_A^+\rangle \langle \Phi_A^+ | A_m^\dagger) \otimes (B_m |\psi\rangle \langle \psi | B_m^\dagger) \xrightarrow{\text{Tr}_{A^{(c)}}} \sum_m s_m^2 B_m |\psi\rangle \langle \psi | B_m^\dagger, \quad (\text{B7})$$

describing a statistical mixture of the states of Eq. (B6) with probabilities  $p_k = \text{Tr}(P_k U(|\Phi_A^+\rangle \langle \Phi_A^+ | \otimes |\psi\rangle \langle \psi |) U^\dagger P_k) = s_k^2 \|B_k |\psi\rangle\|^2$ . Note that from unitarity of  $U$ , we have  $\|U(|\Phi_A^+\rangle \otimes |\psi\rangle)\|^2 = \|\Phi_A^+\rangle \otimes |\psi\rangle\|^2 = 1$  and therefore

$$1 = \|U(|\Phi_A^+\rangle \otimes |\psi\rangle)\|^2 = \sum_{k,l=1}^R s_k s_l \frac{1}{d_A} \langle A_k | A_l \rangle \langle \psi | B_k^\dagger B_l | \psi \rangle = \sum_{k=1}^R s_k^2 \|B_k |\psi\rangle\|^2 = \sum_{k=1}^R p_k. \quad (\text{B8})$$

### APPENDIX C: OPERATOR APPROXIMATIONS USING THE SINGULAR-VALUE DECOMPOSITION

To show that a collection of the  $A_k$  and  $B_k$  also yield an optimal low-rank approximation, we need to show that they form a global minimum of Eq. (5). The approximation of  $U$  by a rank- $k$  tensor decomposition is equivalent to finding a rank- $k$  approximation of  $\mathcal{R}(U)$  from the van Loan and Pitsianis algorithm, i.e.,

$$\left\| \mathcal{R}(U) - \sum_k^r t_k \text{vec}(C_k) \text{vec}(D_k)^\dagger \right\|. \quad (\text{C1})$$

We will show that  $t_k = s_k$ ,  $C_k = A_k$ , and  $D_k = B_k \forall k$  minimizes Eq. (C1). This statement has been proven in Ref. [46] using a distance induced by the spectral norm  $\|\cdot\|_\infty$ . The generalization to the Frobenius distance is well known but proofs are often omitted in the literature. We present one here.

#### 1. Optimal low-rank approximation

*Proposition 1.* Let  $T \in \mathcal{L}(\mathcal{H})$  with a singular-value decomposition  $T = \sum_{i=1}^R \sigma_i u_i v_i^\dagger$ . Let  $T_r$  be the truncation of  $T$  to its  $r$  largest singular values, i.e.,  $T_r = \sum_{i=1}^r \sigma_i u_i v_i^\dagger$ . Then,

$$\min_{\text{rank}(S)=r} \|T - S\| = \|T - T_r\| = \sqrt{\sum_{j=r+1}^R \sigma_j^2}. \quad (\text{C2})$$

*Proof.* The second equality follows directly from the definition of the Frobenius norm. To show the first equality, consider an arbitrary rank- $r$  operator,  $S$ , and calculate the Frobenius distance

$$\|T - S\|^2 = \sum_{i=1}^d \sigma_i (T - S)^2 \geq \sum_{i=1}^{d-r} \sigma_i (T - S)^2, \quad (\text{C3})$$

where we have denoted the  $i$ th singular value of matrix  $A$  by  $\sigma_i(A)$  and dropped the  $r$  smallest singular values to estimate a lower bound. Let  $T_{r+i-2}$  be the rank- $(r+i-2)$  approximation as defined above. We have

$$\sigma_{r+i}(T) = \sigma_1(T - T_{r+i-1}) \leq \sigma_1(T - (T - S)_{i-1} - S_r). \quad (\text{C4})$$

Here, we have denoted the rank- $k$  approximation of  $T$  by  $T_k$ . The inequality follows from the fact that the rank of  $(T - S)_{i-1} + S_r$  is smaller or equal to the rank of  $T_{r+i-1}$ . Next, consider the inequality

$$\sigma_1(A + B) \leq \sigma_1(A) + \sigma_1(B), \quad (\text{C5})$$

which is a direct consequence of the triangular inequality of the spectral norm. With this, we have

$$\begin{aligned} \sigma_1(T - (T - S)_{i-1} - S_r) &\leq \sigma_1(T - S - (T - S)_{i-1}) + \sigma_1(S - S_r) \\ &= \sigma_i(T - S) + \sigma_{r+1}(S). \end{aligned} \quad (\text{C6})$$

Since  $S$  is rank  $r$ , we know that  $\sigma_{r+1}(S) = 0$  and we are left with the overall inequality

$$\sigma_{r+i}(T) \leq \sigma_i(T - S). \quad (\text{C7})$$

Returning to Eq. (C3), we can estimate

$$\|T - S\|^2 \geq \sum_{i=1}^{d-r} \sigma_{r+i}(T)^2 = \sum_{i=r+1}^R \sigma_i(T)^2. \quad (\text{C8})$$

Taking the square root of both sides shows the statement.  $\blacksquare$

## 2. Nearest unitary approximation

In the extreme case in which  $U$  is close to a rank-1 operator, i.e.,  $s_1 \approx 1$  and  $s_k \approx 0$  for  $k \neq 1$ , we expect the dominant product operator  $A_1 \otimes B_1$  to be close to a unitary. This means that  $A_1$  and  $B_1$  have to be close to unitaries  $U_A$  and  $U_B$ , individually. In the following, we find the closest unitaries  $U_A, U_B$  to  $A_1$  and  $B_1$ . This is a well-known problem that is solved by setting all singular values to 1. The following proposition holds for arbitrary unitary-equivalent norms [47]. For pedagogical reasons, let us review the proof for the 2-norm.

*Proposition 2.* Let  $T \in \mathcal{L}(\mathcal{H})$ , with a singular-value decomposition  $T = U\Sigma V^\dagger$ . Then,

$$\min_{W^\dagger W = \mathbb{1}} \|T - W\| = \|T - UV^\dagger\|. \quad (\text{C9})$$

*Proof.* From unitary invariance of the 2-norm, we can reformulate the minimization problem as follows:

$$\min_{W^\dagger W = \mathbb{1}} \|U\Sigma V^\dagger - W\| = \min_{W^\dagger W = \mathbb{1}} \|\Sigma - U^\dagger W V\| = \min_{W^\dagger W = \mathbb{1}} \|\Sigma - W\|, \quad (\text{C10})$$

where we have redefined the unitary  $W$  in the last step using the closedness of the unitary group. The 2-norm can be calculated explicitly as

$$\begin{aligned} \|\Sigma - W\|^2 &= \text{Tr}(\Sigma^2) + \text{Tr}(W^\dagger W) - 2 \text{Re}[\text{Tr}(\Sigma W)] \\ &= d + \sum_k^d (\sigma_k^2 - 2\sigma_k \text{Re}[W_{kk}]), \end{aligned} \quad (\text{C11})$$

where we have denoted the singular values by  $\Sigma_{kl} = \sigma_k \delta_{kl}$ . From unitarity of  $W$ , we know that  $\text{Re}[W_{kk}] \leq 1$  and thus

$$\|\Sigma - W\|^2 \geq d + \sum_k^d \sigma_k^2 - 2\sigma_k = \sum_k^d (\sigma_k - 1)^2 = \|\Sigma - \mathbb{1}\|^2 \forall W. \quad (\text{C12})$$

Therefore, the closest unitary to  $T$  is  $UV^\dagger$ .  $\blacksquare$

The nearest unitary approximation can be used to find a fast quantum transform  $U_{A^{(1)}} \otimes U_{A^{(2)}} \otimes \cdots \otimes U_{A^{(M)}}$  that approximates the action of  $U$ . For a fast quantum transform, we need to iterate the unitary approximation for  $M$  tensor-product factors. Doing so, we introduce two types of errors, one by a rank-1 approximation (cf. Proposition 1) and one by the nearest unitary approximation of the dominant components  $A_1$  (cf. Proposition 2).

*Proposition 3.* Let  $U = \sum_{j_1, \dots, j_M}^{R_1, \dots, R_M} s_{j_1 \dots j_M} A_{j_1}^{(1)} \otimes \cdots \otimes A_{j_M}^{(M)}$  be the multipartite tensor-product decomposition of a unitary  $U$ , with normalized  $\|A_{j_m}^{(m)}\| = \sqrt{d_{A^{(m)}}} \forall m \in \{1, \dots, M\}$ . Further, let  $U_m = V_m W_m^\dagger$  be the nearest unitary approximation of  $A_1^{(m)} = V_m \Sigma_m W_m^\dagger$  (cf. Proposition 2) and  $(1/\sqrt{2d_{A^{(m)}}}) \|A_1^{(m)} - U_m\| = \varepsilon^{(m)} > 0$ , as well as  $1 - s_{1\dots 1} = \varepsilon_s > 0$ . Then,

$$\frac{1}{\sqrt{2d}} \|U - U_1 \otimes \cdots \otimes U_M\| \leq \sqrt{\varepsilon_s} + \sqrt{\frac{1}{2} \varepsilon_s^2 + \sum_{m=1}^M (\varepsilon^{(m)})^2}. \quad (\text{C13})$$



*Proof.* We begin with splitting the error using the triangular inequality:

$$\|U - U_1 \otimes \cdots \otimes U_M\| \leq \left\| s_{1\dots 1} A_1^{(1)} \otimes \cdots \otimes A_1^{(M)} - U_1 \otimes \cdots \otimes U_M \right\| + \left\| \sum_{j_1, \dots, j_M \neq (1, \dots, 1)}^{R_1, \dots, R_M} s_{j_1 \dots j_M} A_{j_1}^{(1)} \otimes \cdots \otimes A_{j_M}^{(M)} \right\|. \quad (\text{C14})$$

Let us consider the two terms separately. First, observe that, by construction,

$$\langle A_1^{(m)} | U_m \rangle_{HS} = \text{Tr} \left( A_1^{(m)\dagger} U_m \right) = \text{Tr} \left( W_m \Sigma W_m^\dagger \right) = \text{Tr} \left( \Sigma \right) = \sum_k^{d_{A_1}} \sigma_k \quad \forall m \in \{1, \dots, M\}, \quad (\text{C15})$$

which is real, i.e.,  $\langle A_1^{(m)} | U_m \rangle_{HS} = \langle U_m | A_1^{(m)} \rangle_{HS}$ . Using this, we can relate the Hilbert-Schmidt product above to the 2-norm error  $\|P - Q\|^2 = \|P\|^2 + \|Q\|^2 - 2 \text{Re} \left( \langle P | Q \rangle_{HS} \right)$  for any two operators  $P, Q$ . With this, we have

$$\begin{aligned} \frac{1}{2d} \left\| s_{1\dots 1} A_1^{(1)} \otimes \cdots \otimes A_1^{(M)} - U_1 \otimes \cdots \otimes U_M \right\|^2 &= \frac{1}{2} \left( 1 + s_{1\dots 1}^2 - 2 \frac{s_{1\dots 1}}{d} \prod_m \langle A_1^{(m)} | U_m \rangle_{HS} \right) \\ &= \frac{1 + s_{1\dots 1}^2}{2} - s_{1\dots 1} \prod_m \left( 1 - \frac{1}{2d_{A^{(m)}}} \left\| A_1^{(m)} - U_m \right\|^2 \right). \end{aligned} \quad (\text{C16})$$

The function  $\prod_m (1 - x_m)$  is convex in the domain  $x_m \in [0, 1] \forall m$ , so we can estimate  $\prod_m (1 - x_m) \geq 1 - \sum_m x_m$  and arrive at

$$\begin{aligned} \frac{1}{2d} \left\| s_{1\dots 1} A_1^{(1)} \otimes \cdots \otimes A_1^{(M)} - U_1 \otimes \cdots \otimes U_M \right\|^2 &\leq \frac{1}{2} (1 - s_{1\dots 1})^2 + \sum_m \frac{1}{2d_{A^{(m)}}} \left\| A_1^{(m)} - U_m \right\|^2 \\ &\leq \frac{1}{2} \varepsilon_s^2 + \sum_m (\varepsilon^{(m)})^2, \end{aligned} \quad (\text{C17})$$

where we have used  $1 - s_{1\dots 1} = \varepsilon_s$ . For the second term of Eq. (C14), we use  $1 - s_{1\dots 1}^2 = 2\varepsilon_s - \varepsilon_s^2 \leq 2\varepsilon_s$ , which also follows from convexity. Using the orthogonality of the  $A_{j_m}^{(m)}$  for fixed  $m$ , the second term reads

$$\frac{1}{2d} \left\| \sum_{j_1, \dots, j_M \neq (1, \dots, 1)}^{R_1, \dots, R_M} s_{j_1 \dots j_M} A_{j_1}^{(1)} \otimes \cdots \otimes A_{j_M}^{(M)} \right\|^2 = \frac{1}{2} \sum_{j_1, \dots, j_M \neq (1, \dots, 1)}^{R_1, \dots, R_M} s_{j_1 \dots j_M}^2 = \frac{1 - s_{1\dots 1}^2}{2} \leq \varepsilon_s, \quad (\text{C18})$$

where the second equality makes use of the normalization of the  $s_k$  and the last inequality is convexity again. This concludes the proof.  $\blacksquare$

## APPENDIX D: ERROR PROPAGATION FOR QTPD

In this appendix, we follow the error coming from tomography throughout QTPD and its applications and herewith give faithful bounds on the error of predictions given a fixed sample budget for tomography.

### 1. Error on tomography and distillation

The  $A_k$  and  $s_k$  are captured from the density matrix  $\rho_A(U)$  [cf. Equation (3)] via diagonalization,

$$\rho_A(U) = \sum_k s_k^2 \text{vec}(A_k) \text{vec}(A_k)^\dagger = V D V^\dagger, \quad (\text{D1})$$

finding the eigenbasis  $V_{ij} = \text{vec}(A_j)_i$  and the eigenvalues  $D_{ij} = s_j^2 \delta_{ij}$ . From the orthogonality and completeness of the  $A_k$ , we can show that  $V$  is unitary, i.e.,  $(V^\dagger V)_{ij} = (1/d_A) \langle A_i | A_j \rangle = \delta_{ij}$ . To get a classical snapshot of Eq. (D1), a tomography is necessary. State tomography suffers from an error that scales inversely with the number  $N_S$  of samples used,

$\|\varepsilon^{(T)}\| = \mathcal{O}\left(\sqrt{d_A^2/N_S}\right)$  [4,23]. Subsequent refinements require only a sample number of  $\mathcal{O}\left(\text{rank}(\rho)(d_A^2/\varepsilon^2)\right)$  [48]—or  $\mathcal{O}\left(d_A^2/\varepsilon^2\right)$ , allowing for a small failure probability [49]. A recent improvement has found that it is necessary to use at least  $\Omega\left(\text{rank}(\rho)d_A^2/\varepsilon\right)$  measurements and it has also been conjectured to be sufficient [50].

Throughout this paper, we consider the 2-norm  $\|\cdot\|$ , which gives an average-case error if divided by the square root of the Hilbert-space dimension. The following discussion can be done straightforwardly for the operator norm, which gives a measure of the worst-case error instead, if dimension factors are correctly accounted for. The difference operator  $\varepsilon^{(T)}$  between the true reduced density matrix  $\rho_A(U)$  and the output of the tomography,

$$\rho_A(U)^{(T)} = \sum_k s_k^{(T)2} \text{vec}(A_k^{(T)})\text{vec}(A_k^{(T)})^\dagger = V^{(T)}D^{(T)}V^{(T)\dagger}, \quad (\text{D2})$$

can then be decomposed into deviation of eigenvalues  $\varepsilon^{(D)}$  and drift of eigenstates  $\varepsilon^{(V)}$  in the following way:

$$\rho_A(U) = \rho_A(U)^{(T)} + \varepsilon^{(T)} = V^{(T)}D^{(T)}V^{(T)\dagger} + \varepsilon^{(T)} = (V - \varepsilon^{(V)})(D - \varepsilon^{(D)})(V - \varepsilon^{(V)})^\dagger + \varepsilon^{(T)}. \quad (\text{D3})$$

Solving for  $\varepsilon^{(T)}$  gives us a relation between those errors. For the sake of a simple presentation, we will estimate the different errors by their maximum  $\varepsilon = \max(\|\varepsilon^{(D)}\|, \|\varepsilon^{(V)}\|)$ . Then,

$$\begin{aligned} \varepsilon^{(T)} &= \varepsilon^{(V)}DV^\dagger + VD\varepsilon^{(V)\dagger} + V\varepsilon^{(D)}V^\dagger + \mathcal{O}(\varepsilon^2), \\ \|\varepsilon^{(T)}\| &= \|\varepsilon^{(V)}DV^\dagger + VD\varepsilon^{(V)\dagger} + V\varepsilon^{(D)}V^\dagger\| + \mathcal{O}(\varepsilon^2) \leq \|\varepsilon^{(V)}D\| + \|VD\varepsilon^{(V)\dagger}V\| + \|V\varepsilon^{(D)}\| + \mathcal{O}(\varepsilon^2) \leq 3\varepsilon + \mathcal{O}(\varepsilon^2). \end{aligned} \quad (\text{D4})$$

In the second-to-last step, we have used the triangle inequality and in the last step, we have used unitary invariance of the 2-norm, as well as submultiplicativity of the 2-norm and  $\|D\| = 1$ . At the end of the day, we are interested in the Schmidt values,  $s_k$ , and the operators  $A_k$  and their predictions,  $s_k^{(T)}$  and  $A_k^{(T)}$ , from  $\rho_A(U)^{(T)}$ . Define

$$\varepsilon_k^{(S)} = s_k^2 - s_k^{(T)2}, \quad \varepsilon_k^{(A)} = A_k - A_k^{(T)}, \quad (\text{D5})$$

where the  $\varepsilon_k^{(S)}$  are scalars and the  $\varepsilon_k^{(A)}$  are operators. The index  $k$  runs through  $\{1, \dots, R^{(T)}\}$  with the tomographic estimate  $R^{(T)}$  of the rank  $R$ . In the most naive scenario,  $R^{(T)}$  will be close or equal to its maximum  $d_A^2$ , as every error  $\varepsilon_k^{(S)} \neq 0$ . Typically, one needs to define a threshold (dependent on  $N_S$ ) below which the  $s_k^{(T)}$  are considered zero. We can further relate eigenvalue deviation to the error of the  $s_k^{(T)}$ ,

$$\|\varepsilon^{(D)}\| = \|D - D^{(T)}\| = \sqrt{\sum_k \left(\varepsilon_k^{(S)}\right)^2} \leq d_A \varepsilon^{(S)}, \quad (\text{D6})$$

where we have defined  $\varepsilon^{(S)} = \max_k \left(\varepsilon_k^{(S)}\right)$ . Similarly, we can relate the eigenstate drift error to the error of the  $A_k^{(T)}$ . Per construction, the  $A_k^{(T)}$  are normalized to  $d_A$  and orthogonal to each other but admit drift angles that are linear in the operators  $\varepsilon_k^{(A)}$ ; to be precise,  $\langle A_k^{(T)} | A_l \rangle = \delta_{kl}d_A + \langle A_k^{(T)} | \varepsilon_l^{(A)} \rangle$ . We collect those drift angles in the matrix  $\varepsilon_{jk}^{(A)} := (\langle A_j^{(T)} | \varepsilon_k^{(A)} \rangle / d_A)$ . In general, all entries of  $\varepsilon^{(A)}$  can be nonzero and of the same order of magnitude:

$$\|\varepsilon_k^{(A)}\| = \|A_k - A_k^{(T)}\| = \sqrt{2d_A} \sqrt{1 - \frac{1}{d_A} \text{Re}[\langle A_k^{(T)} | A_k \rangle]} = \sqrt{-2d_A \text{Re}[\varepsilon_{kk}^{(A)}]}, \quad (\text{D7})$$

$$\|\varepsilon^{(V)}\| = \|V - V^{(T)}\| = \sqrt{2}d_A \sqrt{1 - \frac{1}{d_A^2} \sum_{k,l} \text{Re}[\langle A_k^{(T)} | A_l \rangle]} = \sqrt{-2 \sum_{k,l} \text{Re}[\varepsilon_{kl}^{(A)}]}. \quad (\text{D8})$$

Note that the factor  $d_A$  makes up for the scaling of the 2-norm in Hilbert-space dimension, while the drift-angle matrix  $\varepsilon_{jk}^{(A)}$  does not. The fact that the errors  $\|\varepsilon_k^{(A)}\|$  and  $\|\varepsilon_k^{(V)}\|$  involve the real part  $\text{Re}$  is due to the sensitivity of the norm-induced distance measure to global phases. Since a global phase difference, e.g.,  $D \rightarrow e^{i\varphi}D$ , does not change the outcomes,

we might exchange  $-\text{Re} \left[ \varepsilon_{kl}^{(A)} \right]$  by  $\left| \varepsilon_{kl}^{(A)} \right|$ , which takes the minimum over  $\phi$ , without loss of generality. The faulty  $A_k^{(T)}$  are further used to filter out the action of the  $B_k$ . Instead of measuring the projector  $P_k$  from Eq. (4), we have to use  $P_k^{(T)} = A_k^{(T)} |\Phi_A^+\rangle \langle \Phi_A^+| A_k^{(T)\dagger}$  and obtain the measurement output

$$\begin{aligned} P_k^{(T)} U |\Phi_A^+\rangle |\psi\rangle &= \sum_l^R s_l \frac{\langle A_k^{(T)} | A_l \rangle}{d_A} A_k^{(T)} |\Phi_A^+\rangle \otimes B_l |\psi\rangle = \sum_l^R \left( \delta_{kl} + \frac{\langle A_k^{(T)} | \varepsilon_l^{(A)} \rangle}{d_A} \right) s_l A_k^{(T)} |\Phi_A^+\rangle \otimes B_l |\psi\rangle \\ &= s_k A_k^{(T)} |\Phi_A^+\rangle \otimes B_k |\psi\rangle + \sum_l^R \varepsilon_{kl}^{(A)} s_l A_k^{(T)} |\Phi_A^+\rangle \otimes B_l |\psi\rangle. \end{aligned} \quad (\text{D9})$$

In order to normalize this state, we have to multiply it by

$$\begin{aligned} N(\varepsilon^{(A)}) &:= \frac{1}{s_k \sqrt{d_A}} \left( \sum_{l,m} \lambda_{lm} \left( \delta_{kl} + \frac{s_l}{s_k} \varepsilon_{kl}^{(A)} \right) \left( \delta_{km} + \frac{s_m}{s_k} \varepsilon_{km}^{(A)*} \right) \right)^{-\frac{1}{2}} \\ &= \frac{1}{s_k \sqrt{d_A}} \left( \frac{1}{\sqrt{\lambda_{kk}}} - \frac{1}{\sqrt{\lambda_{kk}}} \sum_l \frac{s_l}{s_k} \text{Re} \left( \frac{\lambda_{lk}}{\lambda_{kk}} \varepsilon_{kl}^{(A)} \right) \right) + \mathcal{O}(\varepsilon^2), \end{aligned} \quad (\text{D10})$$

where we have used the coefficients  $\lambda_{kl} = s_k s_l \langle \psi | B_l^\dagger B_k | \psi \rangle$  from Eq. (20). The difference between the normalized state vectors then reads

$$\begin{aligned} &\left( N(\varepsilon^{(A)}) P_k^{(T)} - N(0) P_k \right) (\mathbb{1}_A \otimes U) (|\Phi_A^+\rangle \otimes |\psi\rangle) \\ &= -\frac{\varepsilon_k^{(A)} |\Phi_A^+\rangle}{\sqrt{d_A}} \otimes \frac{B_k |\psi\rangle}{\sqrt{\lambda_{kk}}} + \frac{A_k^{(T)} |\Phi_A^+\rangle}{\sqrt{d_A}} \otimes \sum_l \sqrt{\frac{\lambda_{ll}}{\lambda_{kk}}} \left( \frac{\varepsilon_{kl}^{(A)} s_l}{s_k} \frac{B_l |\psi\rangle}{\sqrt{\lambda_{ll}}} - \text{Re} \left( \frac{\varepsilon_{kl}^{(A)} s_l}{s_k} \frac{\lambda_{kl}}{\sqrt{\lambda_{kk} \lambda_{ll}}} \right) \frac{B_k |\psi\rangle}{\sqrt{\lambda_{kk}}} \right) + \mathcal{O}(\varepsilon^2). \end{aligned} \quad (\text{D11})$$

To obtain an error measure for the output state of Eq. (B5), we calculate the distance between the normalized states:

$$\begin{aligned} &\left\| \left( N(\varepsilon^{(A)}) P_k^{(T)} - N(0) P_k \right) (\mathbb{1}_A \otimes U) (|\Phi_A^+\rangle \otimes |\psi\rangle) \right\| \\ &\leq \frac{\|\varepsilon_k^{(A)}\|}{\sqrt{d_A}} + \left\| \sum_l \sqrt{\frac{\lambda_{ll}}{\lambda_{kk}}} \left( \frac{\varepsilon_{kl}^{(A)} s_l}{s_k} \frac{B_l |\psi\rangle}{\sqrt{\lambda_{ll}}} - \text{Re} \left( \frac{\varepsilon_{kl}^{(A)} s_l}{s_k} \frac{\lambda_{kl}}{\sqrt{\lambda_{kk} \lambda_{ll}}} \right) \frac{B_k |\psi\rangle}{\sqrt{\lambda_{kk}}} \right) \right\| + \mathcal{O}(\varepsilon^2). \end{aligned} \quad (\text{D12})$$

If we assume, for the sake of simplicity, that the matrix of drift angles is diagonal, i.e.,  $\varepsilon_{jk}^{(A)} = \varepsilon_{kk}^{(A)} \delta_{jk}$ , then the error bound reads

$$\begin{aligned} &\left\| \left( N(\varepsilon^{(A)}) P_k^{(T)} - N(0) P_k \right) (\mathbb{1}_A \otimes U) (|\Phi_A^+\rangle \otimes |\psi\rangle) \right\| \leq \frac{\|\varepsilon_k^{(A)}\|}{\sqrt{d_A}} + \left\| \text{Im} \left( \frac{\varepsilon_{kk}^{(A)}}{\sqrt{\lambda_{kk}}} \right) \frac{B_k |\psi\rangle}{\sqrt{\lambda_{kk}}} \right\| + \mathcal{O}(\varepsilon^2) \\ &\leq \frac{\|\varepsilon_k^{(A)}\|}{\sqrt{d_A}} + \left| \varepsilon_{kk}^{(A)} \right| + \mathcal{O}(\varepsilon^2) = \left( 1 + \frac{1}{\sqrt{2}} \right) \frac{1}{\sqrt{d_A}} \|\varepsilon_k^{(A)}\| + \mathcal{O}(\varepsilon^2), \end{aligned} \quad (\text{D13})$$

where we have used Eq. (D7). The same result can be achieved by neglecting the terms  $s_l/s_k$  for  $l \neq k$ , which is valid as long as  $s_k$  is a dominant singular value. In summary, the error  $\varepsilon^{(T)}$  from tomography propagates linearly (in leading order) through the diagonalization into  $s_k$  and  $A_k$ , as well as through the distillation of the  $B_k$ , and can be suppressed with raising the number of shots  $N_S$ .

## 2. Worst-case error

The 2-norm is a natural choice to measure distances on the space of vectorized operators, as it is induced from the Hilbert-Schmidt inner product. As we have normalized the 2-norm before by a factor of dimension, it represents typical errors. In the following, we leave a short note on the worst-case error, which is measured by the operator norm instead. In analogy to Eq. (D4), we can relate the operator-norm error of the tomography to eigenvalue and drift errors,

$$\|\varepsilon^{(T)}\|_\infty \leq 3\varepsilon_\infty + \mathcal{O}(\varepsilon^2), \quad (\text{D14})$$

where  $\varepsilon_\infty = \max\{\|\varepsilon^{(V)}\|_\infty, \|\varepsilon^{(D)}\|_\infty\}$ , since the operator norm is also unitary invariant and  $\|D\|_\infty \leq 1$ . We can further relate

$$\|\varepsilon^{(D)}\|_\infty = \max_k \left| \varepsilon_k^{(S)} \right|, \quad (\text{D15})$$

$$\|\varepsilon^{(V)}\|_\infty = \|\varepsilon^{(A)}\|_\infty, \quad (\text{D16})$$

where we have defined the matrix  $\varepsilon^{(A)}$  with elements  $\varepsilon_{jk}^{(A)}$ . The second identity follows from unitary invariance. Using the calculation from Eq. (D12), we can straightforwardly bound the error on the distillation of the  $B_k$  via

$$\left\| \left( N(\varepsilon^{(A)})P_k^{(T)} - N(0)P_k \right) U|\Phi_A^+\rangle|\psi\rangle \right\| \leq \frac{\|\varepsilon_k^{(A)}\|}{\sqrt{d_A}} + R \max_l \left( \sqrt{\frac{\lambda_{ll} s_l}{\lambda_{kk} s_k}} \right) \|\varepsilon_{kl}^{(A)}\|_\infty + \mathcal{O}(\varepsilon^2), \quad (\text{D17})$$

which is loose by a factor of  $R \leq d_A^2$ , in general, but is reduced in cases in which the drift is approximately diagonal or only few  $s_k$  are dominant. Also here, the tomography error propagates linearly through the errors for  $s_k, A_k$ , and  $B_k$  and can be suppressed with the number of shots  $N_S$ .

## 3. Error on applications

As for the applications of QTPD, how the error propagates depends on the objective of interest. Let us start with the nonlocality  $S_A(U) = -\sum_k s_k^2 \log s_k^2$ , which only depends on the Schmidt values,  $s_k$ , and thus the error depends only on  $\varepsilon^{(S)} = (\|\varepsilon^{(D)}\|/d_A)$ :

$$\begin{aligned} \left| S_A^{(T)}(U) - S_A(U) \right| &= \left| \sum_k \left( 2 \left( s_k^{(T)} \right)^2 \log s_k^{(T)} - 2s_k^2 \log s_k \right) \right| = \left| \sum_k \log \frac{s_k^{2s_k^2}}{\left( s_k^{(T)} \right)^{2 \left( s_k^{(T)} \right)^2}} \right| \\ &= \left| \sum_k \log \frac{\left( \left( s_k^{(T)} \right)^2 + \varepsilon_k^{(S)} \right)^{\left( s_k^{(T)} \right)^2 + \varepsilon_k^{(S)}}}{\left( s_k^{(T)} \right)^{2 \left( s_k^{(T)} \right)^2}} \right| \\ &= \sum_k \text{sgn}(\varepsilon_k^{(S)}) \left( 1 + \left( s_k^{(T)} \right)^2 \log \left( s_k^{(T)} \right)^2 \right) \varepsilon_k^{(S)} + \mathcal{O} \left( \left( \varepsilon^{(S)} \right)^2 \right). \end{aligned} \quad (\text{D18})$$

In the last step, we have linearized the logarithm in  $\varepsilon_k^{(S)}$ . Since the entangling power from Eq. (E2) is just a sum of nonlocality measures, the error propagates similarly. If we use QTPD for open quantum dynamics, the tomography error also propagates into expectation values of mixed states,  $\sigma_A$ . For errors of expectation values, it is sufficient to consider the trace distance of the reduced density matrices  $T(\sigma_A, \sigma_A^{(T)})$ , since it upper bounds errors in expectation values of observables [51].



*Lemma 4.* Let  $\rho$  and  $\sigma$  be density matrices. The difference between the expectation values of an observable  $O$  can be bounded in the following way:

$$|\mathrm{Tr}(O\rho) - \mathrm{Tr}(O\sigma)| \leq \|O\|_\infty \|\rho - \sigma\|_1. \quad (\text{D19})$$

We are thus left with the 1-norm-induced distance of the reduced density matrix  $\sigma_A^{(T)} = \sum_{k,l} \lambda_{kl} A_k^{(T)} |\psi_A\rangle\langle\psi_A| A_l^{(T)\dagger}$  from  $\sigma_A$  [cf. Eq. (20)]. The difference operator reads

$$\begin{aligned} \sigma_A - \sigma_A^{(T)} &= \sum_{k,l} \lambda_{kl} \left[ \left( A_k^{(T)} + \varepsilon_k^{(A)} \right) |\psi_A\rangle\langle\psi_A| \left( A_l^{(T)\dagger} + \varepsilon_l^{(T)\dagger} \right) - A_k^{(T)} |\psi_A\rangle\langle\psi_A| A_l^{(T)\dagger} \right] \\ &= \sum_{k,l} \lambda_{kl} \left[ \varepsilon_k^{(A)} |\psi_A\rangle\langle\psi_A| A_l^{(T)\dagger} + A_k^{(T)} |\psi_A\rangle\langle\psi_A| \varepsilon_l^{(A)\dagger} \right] + \mathcal{O}\left(\left(\varepsilon^{(A)}\right)^2\right). \end{aligned} \quad (\text{D20})$$

We have ignored possible errors in the determination of the coefficients  $\lambda_{kl}$ . Finally, we can estimate an upper bound for the 1-norm error in the following way:

$$\begin{aligned} \left\| \sigma_A - \sigma_A^{(T)} \right\|_1 &\leq \sum_{k,l} |\lambda_{k,l}| \left\| \varepsilon_k^{(A)} |\psi_A\rangle\langle\psi_A| A_l^{(T)\dagger} + A_k^{(T)} |\psi_A\rangle\langle\psi_A| \varepsilon_l^{(A)\dagger} \right\|_1 + \mathcal{O}\left(\left(\varepsilon^{(A)}\right)^2\right) \\ &\leq 2 \sum_{k,l} |\lambda_{k,l}| \left\| \varepsilon_k^{(A)} |\psi_A\rangle\langle\psi_A| A_l^{(T)\dagger} \right\|_1 + \mathcal{O}\left(\left(\varepsilon^{(A)}\right)^2\right) \\ &\leq 2 \sum_{k,l} |\lambda_{k,l}| \left| \langle\psi_A| A_l^{(T)} A_l^{(T)\dagger} |\psi_A\rangle \right| \left\| \varepsilon_k^{(A)} \right\|_2 + \mathcal{O}\left(\left(\varepsilon^{(A)}\right)^2\right) \\ &\leq \left( 2 \sum_{k,l} |\lambda_{k,l}| \left| \langle\psi_A| A_l^{(T)} A_l^{(T)\dagger} |\psi_A\rangle \right| \right) \max_k \varepsilon_k^{(A)} + \mathcal{O}\left(\left(\varepsilon^{(A)}\right)^2\right), \end{aligned} \quad (\text{D21})$$

where we have used the triangle inequality in the first and second steps together with  $\|A\|_1 = \|A^\dagger\|_1$ . In the third step, the Hölder inequality,  $\|AB\|_1 \leq \|A\|_2 \|B\|_2$ , has been used.

## APPENDIX E: ENTANGLEMENT GENERATION FROM QTPD

One measure that singles out the nonlocal but also nonentangling action of the SWAP operator has been introduced in Ref. [6] for the case of equally large subsystems, i.e.,  $d_A = d_B$ :

$$e_A(U) = \frac{1}{\log(d_A^2)} \left( S_A(U) + (S_A(UP_{AB}) - \log(d_A^2)) \right), \quad (\text{E1})$$

where  $P_{AB}$  swaps the two subsystems, i.e.,  $P_{AB}|\psi\rangle_A|\psi\rangle_B = |\psi\rangle_B|\psi\rangle_A$ . Since we are considering the asymmetrical case  $d_A \leq d_B$ , let us define a straightforward generalization in which we sum over all different contributions from permutations between  $A$  and subsystems of dimension  $d_A$  in  $B$ ,

$$e_A(U) = \frac{1}{\log(d_A^2)} \left( S_A(U) + \sum_{\substack{C \subset B \\ \dim(C)=d_A}} (S_A(UP_{AC}) - \log(d_A^2)) \right), \quad (\text{E2})$$

where the sum over subsystems  $C$  only ranges over qubit configurations and is therefore finite. An alternative measure for the entangling power is the mean entanglement that is generated by the action of  $U$  on product states

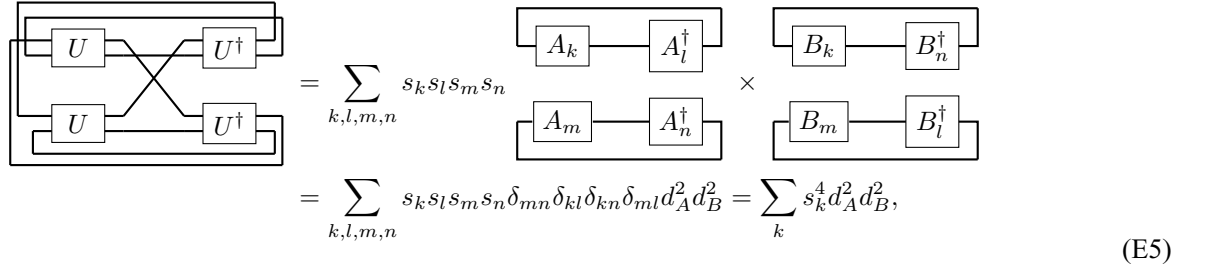
$$e_m(U) = \mathbb{E}_{|\psi_A\rangle, |\psi_B\rangle} [E(\mathrm{Tr}_B(U|\psi_A\rangle \otimes |\psi_B\rangle))], \quad (\text{E3})$$

where  $E$  is an (*a priori* unspecified) measure of entanglement and  $\mathbb{E}$  denotes the Haar measure over the subsystems  $\mathcal{H}_A$  and  $\mathcal{H}_B$ . For the linearized entanglement entropy  $E(\rho) = 1 - \mathrm{Tr}(\rho^2)$ , the entangling power from Eq. (E3) has been

discussed by Zanardi *et al.* [26]. The mean linear entanglement entropy growth from the action of a unitary  $U$  reads [cf. Eq. (5) of [26]]

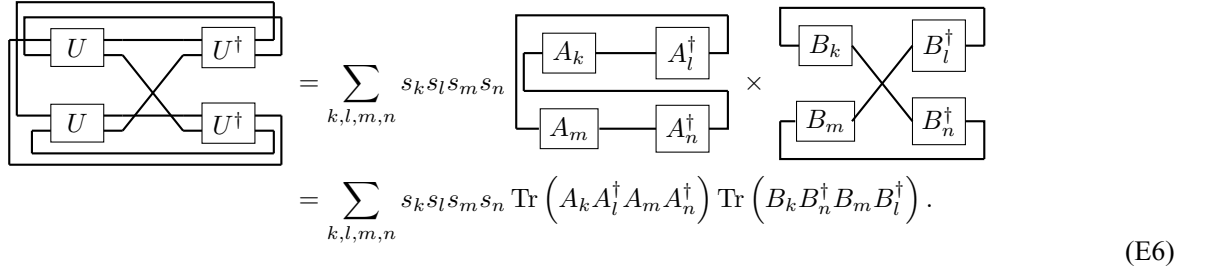
$$e_m(U) = 1 - \frac{1}{d_A(d_A + 1)} \frac{1}{d_B(d_B + 1)} (d_A^2 d_B + d_A d_B^2 + \text{Tr}((U \otimes U) P_A (U^\dagger \otimes U^\dagger) P_A) + \text{Tr}((U \otimes U) P_B (U^\dagger \otimes U^\dagger) P_A)), \quad (\text{E4})$$

using two copies of the full system  $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_A \otimes \mathcal{H}_B$ , where  $P_A |\psi_A\rangle \otimes |\psi_B\rangle \otimes |\phi_A\rangle \otimes |\phi_B\rangle = |\phi_A\rangle \otimes |\psi_B\rangle \otimes |\psi_A\rangle \otimes |\phi_B\rangle$  swaps the states in the two copies of  $\mathcal{H}_A$  and analogously so does  $P_B$  on  $\mathcal{H}_B$ . Note the difference compared to the SWAP operators  $P_{AC}$  that have been used previously. We can simplify the trace terms by inserting  $U = \sum_k s_k A_k \otimes B_k$ . For the first term, we obtain, in tensor-network notation,



$$= \sum_{k,l,m,n} s_k s_l s_m s_n \begin{array}{c} \boxed{A_k} \text{---} \boxed{A_l^\dagger} \\ \boxed{A_m} \text{---} \boxed{A_n^\dagger} \end{array} \times \begin{array}{c} \boxed{B_k} \text{---} \boxed{B_n^\dagger} \\ \boxed{B_m} \text{---} \boxed{B_l^\dagger} \end{array} = \sum_{k,l,m,n} s_k s_l s_m s_n \delta_{mn} \delta_{kl} \delta_{kn} \delta_{ml} d_A^2 d_B^2 = \sum_k s_k^4 d_A^2 d_B^2, \quad (\text{E5})$$

where the second step makes use of the orthonormality of the  $A_k$  and  $B_k$ . The second trace can be calculated analogously:



$$= \sum_{k,l,m,n} s_k s_l s_m s_n \begin{array}{c} \boxed{A_k} \text{---} \boxed{A_l^\dagger} \\ \boxed{A_m} \text{---} \boxed{A_n^\dagger} \end{array} \times \begin{array}{c} \boxed{B_k} \text{---} \boxed{B_l^\dagger} \\ \boxed{B_m} \text{---} \boxed{B_n^\dagger} \end{array} = \sum_{k,l,m,n} s_k s_l s_m s_n \text{Tr} \left( A_k A_l^\dagger A_m A_n^\dagger \right) \text{Tr} \left( B_k B_n^\dagger B_m B_l^\dagger \right). \quad (\text{E6})$$

In summary, the mean entanglement generation reads

$$e_m(U) = 1 - \frac{d_A + d_B}{(d_A + 1)(d_B + 1)} - \frac{d_A d_B \sum_k s_k^4}{(d_A + 1)(d_B + 1)} - \frac{1}{d_A(d_A + 1)} \frac{1}{d_B(d_B + 1)} \sum_{k,l,m,n} s_k s_l s_m s_n \text{Tr} \left( A_k A_l^\dagger A_m A_n^\dagger \right) \text{Tr} \left( B_k B_n^\dagger B_m B_l^\dagger \right). \quad (\text{E7})$$

The last term cannot be calculated without tomographic knowledge of the  $B_k$  and is hence out of reach for near-term quantum computing. It could be solved by a fault-tolerant device. We leave this for future work.

## APPENDIX F: DECOHERENCE-FREE STRUCTURES

To support the discussion on using QTPD for mereology, we give an example for a product operator transformed into an entangled basis and, furthermore, give a characterization of unitaries that admit such a basis in which the action is nonentangling.

### 1. Example for growth of operator entanglement

Consider a two-qubit Hilbert space  $\mathcal{H} \cong \mathcal{H}_A \otimes \mathcal{H}_B$  with a nonentangling T-gate  $U = \mathbb{1} \otimes T$ ,  $T = \begin{bmatrix} 1 & 0 \\ 0 & e^{i(\pi/4)} \end{bmatrix}$ . Further consider the unitarily equivalent gate  $VUV^\dagger$  that is related to  $U$  by a controlled-NOT (CNOT) rotation,  $V = |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes X = V^\dagger$ . One can derive

$$VUV^\dagger = |0\rangle\langle 0| \otimes T + e^{i\frac{\pi}{4}} |1\rangle\langle 1| \otimes T^\dagger. \quad (\text{F1})$$

To show that this is an entangling gate, we can calculate the overlaps with Pauli operators in subsystem  $\mathcal{H}_A$ :

$$\text{Tr}(VUV^\dagger) = T + e^{i\frac{\pi}{4}} T^\dagger = (1 + e^{i\frac{\pi}{4}}) \mathbb{1}, \quad (\text{F2})$$

$$\text{Tr}(Z_A VUV^\dagger) = T - e^{i\frac{\pi}{4}} T^\dagger = (1 - e^{i\frac{\pi}{4}}) Z, \quad (\text{F3})$$

$$\text{Tr}(X_A VUV^\dagger) = 0, \quad (\text{F4})$$

$$\text{Tr}(Y_A VUV^\dagger) = 0. \quad (\text{F5})$$

Altogether, we deduce  $VUV^\dagger = (1 + e^{i(\pi/4)}) \mathbb{1} + (1 - e^{i(\pi/4)}) Z \otimes Z$ .

## 2. A necessary and sufficient condition for the existence of decoherence-free subsystems

*Proposition 5.* Let  $U \in \mathcal{L}(\mathcal{H})$  be a unitary with eigenstates  $|\theta_i\rangle$ , i.e.,  $U|\theta_i\rangle = e^{i\theta_i}|\theta_i\rangle$ . Further, let  $\mathcal{H} \cong \mathcal{H}_A \otimes \mathcal{H}_B$  with  $\dim(\mathcal{H}_A) =: d_A < d_B := \dim(\mathcal{H}_B)$  define a split indexed by  $i = (\mu, m)$ ,  $\mu \in \{1, \dots, d_A\}$ ,  $m \in \{1, \dots, d_B\}$  and  $i \in \{1, \dots, d_A d_B\}$ . Then,

$\exists$  ordering  $\theta_i = \theta_{\mu m} = \phi_\mu + \psi_m$  for some  $\phi_\mu, \psi_m \in [0, 2\pi)$

$$\iff$$

$\exists V$  unitary, such that  $VUV^\dagger = A \otimes B$  with  $A, B$  unitary. (F6)

*Proof.* The backward direction “ $\Leftarrow$ ” becomes trivial as soon as we write down  $U$  in diagonal form. Let  $T$  denote the eigenbasis of  $U$ , i.e.,

$$U = T^\dagger D_U T = T^\dagger \begin{bmatrix} e^{i\theta_1} & & \\ & \ddots & \\ & & e^{i\theta_d} \end{bmatrix} T = V^\dagger (A \otimes B) V$$

$$\iff A \otimes B = VT^\dagger D_U TV. \quad (\text{F7})$$

Thus,  $U$  and  $A \otimes B$  share the same eigenvalues. If we denote the eigenvectors of  $A$  by  $|\phi_\mu\rangle$  and of  $B$  by  $|\psi_m\rangle$ , we obtain  $(A \otimes B)|\phi_\mu\rangle \otimes |\psi_m\rangle = e^{i(\phi_\mu + \psi_m)}|\phi_\mu\rangle \otimes |\psi_m\rangle$  and thus  $\theta_{\mu m} = \phi_\mu + \psi_m$ .

For the forward direction “ $\Rightarrow$ ,” consider again the eigenbasis  $V$  of  $U$ . As we know that the eigenvalues are related by  $\theta_i = \phi_\mu + \psi_m$ , we can put  $V$  into an order such that  $D_U = D_A \otimes D_B$  decouples with

$$D_A = \begin{bmatrix} e^{i\phi_1} & & \\ & \ddots & \\ & & e^{i\phi_{d_A}} \end{bmatrix} \quad \text{and}$$

$$D_B = \begin{bmatrix} e^{i\psi_1} & & \\ & \ddots & \\ & & e^{i\psi_{d_B}} \end{bmatrix}. \quad (\text{F8})$$

We conclude with  $U = V^\dagger (D_A \otimes D_B) V$ . ■

## APPENDIX G: GENERALIZATION TO ARBITRARY OPERATORS

One might be interested in a tensor decomposition of a nonunitary operator  $T \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ , e.g., a Hermitian operator. QTPD can be generalized to general nonunitary operators by utilizing the concept of block encodings [52]. If  $U$  is an  $(\alpha, a, \varepsilon)$  block encoding of  $T$ , we can use the same circuits as in Eqs. (B2) and (B5) together with postselection on  $|0 \dots 0\rangle$  on the ancillary system, i.e.,

$$(G1)$$

The generalization to nonunitary operators  $T$  comes at a price of raising the sample complexity in the two circuits of Eq. (G1). In particular, the sample number will be multiplied by a factor exponential in the number of ancilla qubits  $2^a$ . This puts a restriction onto the tensors  $T$  that can be analyzed this way. While there are typically upper bounds for  $a$  polynomial in the qubit number  $n$  [52], one would need a block encoding with  $a = \mathcal{O}(1)$  in order to keep the sample complexity below full tomography.

## APPENDIX H: ANALYTICAL DISCUSSION OF THE HEISENBERG MODEL ON TWO QUBITS

Consider the Hamiltonian of the Heisenberg model for two qubits and the time-evolution operator  $U$  that we consider as a black-box unitary for tensor decomposition:

$$H = -(J_x X_1 X_2 + J_y Y_1 Y_2 + J_z Z_1 Z_2) \quad \text{and}$$

$$U := e^{-itH} = e^{iJ_x t X_1 X_2} e^{iJ_y t Y_1 Y_2} e^{iJ_z t Z_1 Z_2}. \quad (\text{H1})$$

In the following, we perform all calculations with distinct interaction strengths  $J_x, J_y$  and  $J_z$  and view the isotropic case as an example in which  $J_x = J_y = J_z =: J$ . The time-evolution operator  $U$  decays into local exponentials, because—on two qubits—all terms in the Hamiltonian commute. With the identity for Pauli exponentials  $e^{iJ_x t X_1 X_2} = \cos(J_x t) \mathbb{1} + i \sin(J_x t) X_1 X_2$  (and similar for the other two Pauli strings), we can directly calculate the tensor decomposition of  $U$ :

$$U = (c_x \mathbb{1} + i s_x X_1 X_2) (c_y \mathbb{1} + i s_y Y_1 Y_2) (c_z \mathbb{1} + i s_z Z_1 Z_2)$$

$$= (c_x c_y c_z + i s_x s_y s_z) \mathbb{1} + (i s_x c_y c_z + c_x s_y s_z) X_1 X_2$$

$$+ (i s_y c_x c_z + c_y s_x s_z) Y_1 Y_2 + (i s_z c_y c_x + c_z s_y s_x) Z_1 Z_2$$

$$=: g_0 \mathbb{1} + g_x X_1 X_2 + g_y Y_1 Y_2 + g_z Z_1 Z_2. \quad (\text{H2})$$

We have introduced the shorthand notation  $c_i := \cos(J_i t)$  and  $s_i = \sin(J_i t) \forall i \in \{x, y, z\}$  and omitted the time dependence in the following for the sake of clarity. We also allow ourselves some flexibility in pushing complex phases between  $s_k$  and  $B_k$ , which is technically against our convention introduced in Appendix A, assuming the  $s_k$  to be real. With Eq. (H2), we can read off the Schmidt coefficients  $g_i$  and deduce that  $U$  has maximal Schmidt rank 4 except for when one of the terms vanishes. In the isotropic case, this happens for  $t = (\pi/2J)$ . If we require  $J_z = 0$  and fix  $J_x = J_y = J$ , we recover the SWAP gate  $S$  at  $t = (\pi/4J)$ , i.e.,

$$S = e^{i\frac{\pi}{4}(X_1 X_2 + Y_1 Y_2)} = \frac{1}{2} (\mathbb{1} + X_1 X_2 + Y_1 Y_2 + Z_1 Z_2). \quad (\text{H3})$$

Hence, we obtain the tensor decomposition of the SWAP gate for free from Eq. (H2). We can use this to write down the entangling power of  $U$  on qubit 1:

$$e_1(U) = \frac{1}{\log(4)} (S_1(U) + S_1(US)) - 1, \quad (\text{H4})$$

$$S_1(U) = - \sum_k |g_k|^2 \log |g_k|^2, \quad (\text{H5})$$

$$S_1(US) = -\frac{1}{4} |g_0 + g_x + g_y + g_z|^2 \log \frac{|g_0 + g_x + g_y + g_z|^2}{4} - \frac{1}{4} |g_0 + g_x - g_y - g_z|^2 \log \frac{|g_0 + g_x - g_y - g_z|^2}{4} \\ - \frac{1}{4} |g_0 - g_x + g_y - g_z|^2 \log \frac{|g_0 - g_x + g_y - g_z|^2}{4} - \frac{1}{4} |g_0 - g_x - g_y + g_z|^2 \log \frac{|g_0 - g_x - g_y + g_z|^2}{4}. \quad (\text{H6})$$

One can see directly that for  $U = S$ , the entangling power vanishes while the nonlocality measure  $S_1(U)$  is maximal. In order to obtain a large entangling power, both  $S_1(U)$  and  $S_1(US)$  have to be large. The mean entanglement generation, which is derived in Appendix E, behaves similarly in the two-qubit case. The last term in Eq. (E7) is the only nontrivial term to evaluate. The only nonvanishing trace terms yield

$$\sum_{k,l,m,n} g_k g_l^* g_m g_n^* \text{Tr} \left( A_k A_l^\dagger A_m A_n^\dagger \right) \text{Tr} \left( B_k B_n^\dagger B_m B_l^\dagger \right) = d^2 \sum_k |g_k|^4 + d^2 \sum_{k \neq l} (2 |g_k|^2 |g_l|^2 + g_k^2 g_l^{2*}) \\ + d^2 \sum_{\sigma \in S_4} g_{\sigma(0)} g_{\sigma(x)}^* g_{\sigma(y)} g_{\sigma(z)}^*, \quad (\text{H7})$$

with  $d = 2$ . The three terms arise from different combinations of inserting the Pauli operators  $A_k, B_k \in \{\mathbb{1}, X, Y, Z\}$ . Since Pauli operators are traceless, the products  $A_k A_l^\dagger A_m A_n^\dagger$  have to result in  $\mathbb{1}$ , such that the trace yields a factor of dimension  $d$ . The first term comes from traces of the form  $\text{Tr}(A_k^2)$  and the second from  $\text{Tr}(A_k^2 A_l^2)$ , as well as  $\text{Tr}(A_k A_l A_k A_l)$  with  $k \neq l$  and adequate combinatorial coefficients. Finally, the third term represents traces in which all operators are different, i.e.,  $\text{Tr}(A_k A_l A_m A_n) \text{Tr}(A_k A_n A_m A_l) = -\text{Tr}(A_k A_l A_m A_n)^2$  with  $(k, l, m, n) = (\sigma(0), \sigma(x), \sigma(y), \sigma(z)), \sigma \in S_4$ .

Now, let us fix the initial state to be  $|10\rangle$ . The effective open time evolution of the first qubit under the Hamiltonian  $H$  [Eq. (H1)] can be expressed as a quantum channel  $\mathcal{E}_t$  that evolves a density matrix describing the quantum state of qubit 1:

$$\mathcal{E}_t(|1\rangle\langle 1|) = \rho_1(t) = (|g_0|^2 + |g_z|^2 - g_0 g_z^* - g_z g_0^*) |1\rangle\langle 1| + (|g_x|^2 + |g_y|^2 + g_x g_y^* + g_y g_x^*) |0\rangle\langle 0| \\ = |g_0 - g_z|^2 |1\rangle\langle 1| + |g_x + g_y|^2 |0\rangle\langle 0| \\ = \cos^2((J_x + J_y)t) |1\rangle\langle 1| + \sin^2((J_x + J_y)t) |0\rangle\langle 0|. \quad (\text{H8})$$

We can read off the Schmidt values of this state  $|\lambda_0|^2 = \sin^2((J_x + J_y)t)$  and  $|\lambda_1|^2 = \cos^2((J_x + J_y)t)$ . In order to learn the properties of the output state, we measure the magnetization  $\langle Z \rangle_{\rho_1(t)}$  and the entanglement entropy  $S(\rho_1(t))$  of the time-evolved state in the numerical experiment. In the two-qubit case, we can write down the two observables as functions of



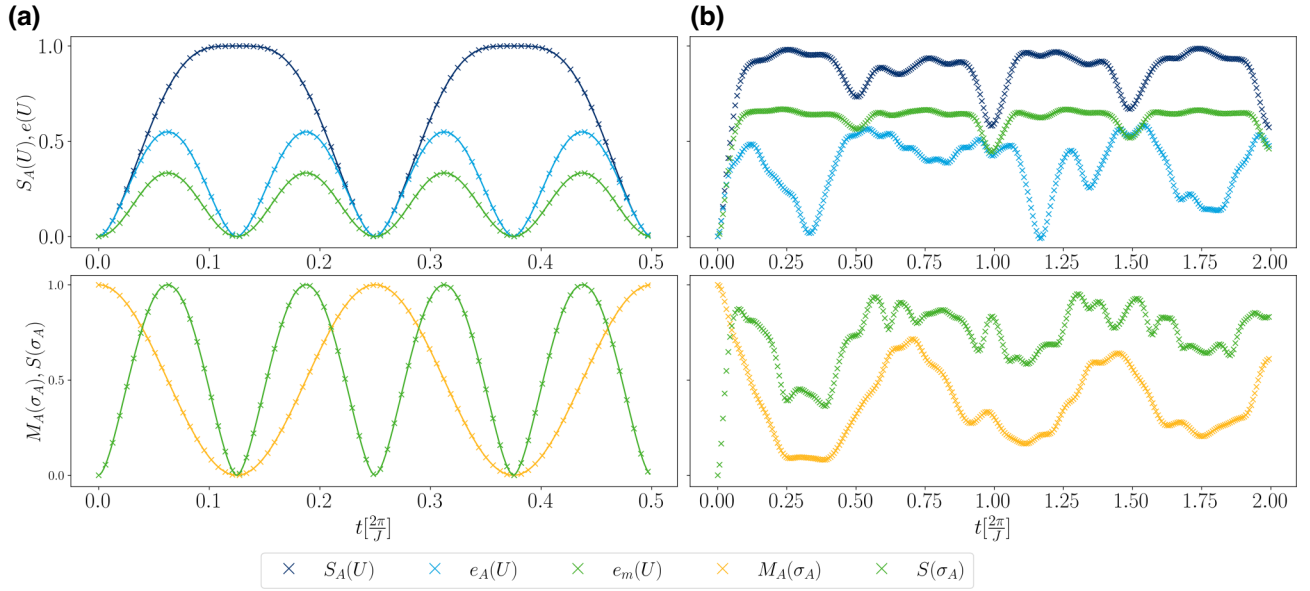


FIG. 4. The open quantum dynamics for the isotropic Heisenberg model of (a) a subsystem of one of two neighboring qubits and (b) two qubits of a 2D grid of  $3 \times 2$  qubits. The upper panel shows the nonlocality measure  $S_A(U)$  and the entangling-power measures  $e_A(U)$  and  $e_m(U)$  (see Appendix E) for the respective time-evolution operator, while the lower panel contains the total magnetization  $M_A(\sigma_A)$  and the entanglement entropy  $S(\sigma_A)$  of the time-evolved state. The initial state of the total system reads  $|1\rangle \otimes |0\rangle$  in (a) and  $|11\rangle \otimes |0000\rangle$  in (b). In (a), the solid curves represent analytical predictions derived in Eqs. (H4)–(H10).

time:

$$\langle Z \rangle_{\rho_1(t)} = \sin^2((J_x + J_y)t) - \cos^2((J_x + J_y)t), \quad (\text{H9})$$

$$S(\rho_1(t)) = \sin^2((J_x + J_y)t) \log(\sin^2((J_x + J_y)t)) + \cos^2((J_x + J_y)t) \log(\cos^2((J_x + J_y)t)). \quad (\text{H10})$$

The above example starts with a product state  $|10\rangle$  and does not show any quantum coherence after time evolution. If we start with a product state  $|1+\rangle$ , which lies skew in two spin-symmetry sectors, some of the coherence on qubit 2 gets swapped to qubit 1:

$$\mathcal{E}_t(|1\rangle\langle 1|) = \rho_1(t) = (|g_0|^2 + |g_z|^2)|1\rangle\langle 1| + (|g_x|^2 + |g_y|^2)|0\rangle\langle 0| + (g_z g_y^* - g_0 g_x^*)|1\rangle\langle 0| + (g_y g_z^* - g_x g_0^*)|0\rangle\langle 1|, \quad (\text{H11})$$

as opposed to Eq. (H8). In Fig. 4, we repeat the numerical experiment shown in the main text but also include the entangling-power measures  $e_A(U)$  and  $e_m(U)$ , introduced in Appendix E. We compare the numerical results with the above analytical derivation and find an exact match. As we have already discussed parts of Fig. 4 in the main text, here we focus on features of the entangling powers.

In the two-qubit case [cf. Fig. 4(a)], the nonlocality undergoes four oscillation periods, representing the oscillation of  $U$  between the identity and the SWAP operator on two qubits. In accordance with this, the entangling powers both show an oscillation and vanish at the extreme points of the nonlocality function  $S_A(U)$ . As they reach zero when  $U$  equals the SWAP operator, while the nonlocality stays maximal, they thus both undergo twice as many oscillations.

On a  $3 \times 2$  qubit grid [cf. Fig. 4(b)], the entangling-power measures no longer follow defined oscillations but start off at zero, quickly rise, and then stay at a nonzero value for most of the time.  $e_A(U)$ , since it singles out SWAP operations, does become close to zero for certain simulation times. However, the trial state shows nonzero entanglement entropy at those times. As a consequence,  $e_A(U)$  does not seem to be a good measure of the entangling power when the dimensions  $d_A$  and  $d_B$  of subsystems  $\mathcal{H}_A$  and  $\mathcal{H}_B$  no longer match. The mean entanglement generation  $e_m(U)$ , on the other hand, is not in disagreement with this. Similar to the nonlocality, it rises quickly but stays around approximately 65% of its maximal value, admitting dips at the same points as the nonlocality  $S_A(U)$ , thus allowing for nonzero entanglement throughout the time interval considered.

- [1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, *Rev. Mod. Phys.* **81**, 865 (2009).
- [2] F. Arute, *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature* **574**, 505 (2019).
- [3] C. F. Van Loan and N. Pitsianis, in *Linear Algebra for Large Scale and Real-Time Applications*, edited by M. S. Moonen, G. H. Golub, and B. L. R. De Moor (Springer Netherlands, Dordrecht, 1993), p. 293.
- [4] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
- [5] W. Dür, G. Vidal, and J. I. Cirac, Optimal conversion of nonlocal unitary operations, *Phys. Rev. Lett.* **89**, 057901 (2002).
- [6] B. Jonnadula, P. Mandayam, K. Życzkowski, and A. Lakshminarayan, Entanglement measures of bipartite quantum gates and their thermalization under arbitrary interaction strength, *Phys. Rev. Res.* **2**, 043126 (2020).
- [7] S. Balakrishnan and R. Sankaranarayanan, Operator-Schmidt decomposition and the geometrical edges of two-qubit gates, *Quantum Inf. Process.* **10**, 449 (2011).
- [8] T. Prosen and I. Pižorn, Operator space entanglement entropy in a transverse Ising chain, *Phys. Rev. A* **76**, 032316 (2007).
- [9] T. Zhou and D. J. Luitz, Operator entanglement entropy of the time evolution operator in chaotic systems, *Phys. Rev. B* **95**, 094206 (2017).
- [10] B. Bertini, P. Kos, and T. Prosen, Operator entanglement in local quantum circuits I: Chaotic dual-unitary circuits, *SciPost Phys.* **8**, 067 (2020).
- [11] C. Zhang, S. Denker, A. Asadian, and O. Gühne, Analyzing quantum entanglement with the Schmidt decomposition in operator space, [arXiv:2304.02447](https://arxiv.org/abs/2304.02447).
- [12] C. F. Van Loan, The ubiquitous Kronecker product, *J. Comput. Appl. Math.* **123**, 85 (2000).
- [13] G. Smith, J. A. Smolin, X. Yuan, Q. Zhao, D. Girolami, and X. Ma, Quantifying coherence and entanglement via simple measurements, [arXiv:1707.09928](https://arxiv.org/abs/1707.09928).
- [14] S. Subramanian and M.-H. Hsieh, Quantum algorithm for estimating  $\alpha$ -Renyi entropies of quantum states, *Phys. Rev. A* **104**, 022428 (2021).
- [15] T. Zhang, G. Smith, J. A. Smolin, L. Liu, X.-J. Peng, Q. Zhao, D. Girolami, X. Ma, X. Yuan, and H. Lu, Quantification of entanglement and coherence with purity detection, [arXiv:2308.07068](https://arxiv.org/abs/2308.07068).
- [16] J. E. Tyson, Operator-Schmidt decompositions and the Fourier transform, with applications to the operator-Schmidt numbers of unitaries, *J. Phys. A: Math. Gen.* **36**, 10101 (2003).
- [17] R. Mansuroglu, T. Eckstein, L. Nützel, S. A. Wilkinson, and M. J. Hartmann, Variational Hamiltonian simulation for translational invariant systems via classical pre-processing, *Quantum Sci. Technol.* **8**, 025006 (2023).
- [18] R. Mansuroglu, F. Fischer, and M. J. Hartmann, Problem-specific classical optimization of Hamiltonian simulation, *Phys. Rev. Res.* **5**, 043035 (2023).
- [19] C. M. Keever and M. Lubasch, Classically optimized Hamiltonian simulation, *Phys. Rev. Res.* **5**, 023146 (2023).
- [20] N. Meyer, D. Scherer, A. Plinge, C. Mutschler, and M. Hartmann, in *Proceedings of the 40th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 202, edited by A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett (PMLR, 2023), p. 24592. <https://proceedings.mlr.press/v202/meyer23a.html>.
- [21] M. Cerezo, M. Larocca, D. García-Martín, N. L. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp, *et al.*, Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing, [arXiv:2312.09121](https://arxiv.org/abs/2312.09121).
- [22] M. Sedláč, A. Bisio, and M. Ziman, Optimal probabilistic storage and retrieval of unitary channels, *Phys. Rev. Lett.* **122**, 170502 (2019).
- [23] J. van Apeldoorn, A. Cornelissen, A. Gilyén, and G. Nannicini, in *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)* (Society for Industrial and Applied Mathematics, Florence, Italy, 2023), p. 1265.
- [24] L. N. Trefethen and D. Bau, *Numerical Linear Algebra* (SIAM, Philadelphia, PA, USA, 1997), Vol. 50.
- [25] N. Halko, P. G. Martinsson, and J. A. Tropp, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, *SIAM Rev.* **53**, 217 (2011).
- [26] P. Zanardi, C. Zalka, and L. Faoro, Entangling power of quantum evolutions, *Phys. Rev. A* **62**, 030301 (2000).
- [27] J. Eisert, Entangling power and quantum circuit complexity, *Phys. Rev. Lett.* **127**, 020501 (2021).
- [28] S. M. Carroll and A. Singh, Quantum mereology: Factorizing Hilbert space into subsystems with quasiclassical dynamics, *Phys. Rev. A* **103**, 022213 (2021).
- [29] M. Tegmark, Consciousness as a state of matter, *Chaos Solit. Fractals* **76**, 238 (2015).
- [30] A. Adil, M. S. Rudolph, A. Arrasmith, Z. Holmes, A. Albrecht, and A. Sornborger, A search for classical subsystems in quantum worlds, [arXiv:2403.10895](https://arxiv.org/abs/2403.10895).
- [31] D. A. Lidar and K. B. Whaley, in *Irreversible Quantum Dynamics* (Springer-Verlag, Berlin, 2003), p. 83.
- [32] R. Mansuroglu and H. Sahlmann, No invariant perfect qubit codes, *J. High Energy Phys.* **2023**, 62 (2023).
- [33] F. Bernards and O. Gühne, Multiparticle singlet states cannot be maximally entangled for the bipartitions, *J. Math. Phys.* **65**, 0159105 (2024).
- [34] B. Pirvu, V. Murg, J. I. Cirac, and F. Verstraete, Matrix product operator representations, *New J. Phys.* **12**, 025012 (2010).
- [35] D. Pérez-García, F. Verstraete, M. M. Wolf, and J. I. Cirac, Matrix product state representations, *Quantum Inf. Comput.* **7**, 401 (2007).
- [36] F. Verstraete, V. Murg, and J. Cirac, Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems, *Adv. Phys.* **57**, 143 (2008).
- [37] A. Eddins, M. Motta, T. P. Gujarati, S. Bravyi, A. Mezzacapo, C. Hadfield, and S. Sheldon, Doubling the size of

- quantum simulators by entanglement forging, *PRX Quantum* **3**, 010309 (2022).
- [38] A. Luongo, Quantum algorithms for data analysis (2023), <https://quantumalgorithms.org/chapter-intro.html#modified-hadamard-test>.
- [39] M. C. Caro, Learning quantum processes and Hamiltonians via the Pauli transfer matrix, *ACM Trans. Quantum Comput.* **5**, 1 (2024).
- [40] M. C. Caro, H.-Y. Huang, N. Ezzell, J. Gibbs, A. T. Sornborger, L. Cincio, P. J. Coles, and Z. Holmes, Out-of-distribution generalization for learning quantum dynamics, *Nat. Commun.* **14**, 3751 (2023).
- [41] M. Nibbi and C. B. Mendl, Block encoding of matrix product operators, [arXiv:2312.08861](https://arxiv.org/abs/2312.08861).
- [42] K. Chen and L.-A. Wu, A matrix realignment method for recognizing entanglement, *Quantum Inf. Comput.* **3** (3), 193 (2003).
- [43] O. Gühne and G. Tóth, Entanglement detection, *Phys. Rep.* **474**, 1 (2009).
- [44] J. Guth Jarkovský, A. Molnár, N. Schuch, and J. I. Cirac, Efficient description of many-body systems with matrix product density operators, *PRX Quantum* **1**, 010304 (2020).
- [45] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, and J. R. McClean, Quantum advantage in learning from experiments, *Science* **376**, 1182 (2022).
- [46] G. Golub and C. Van Loan, in *Matrix Computations*, Johns Hopkins Studies in the Mathematical Sciences (Johns Hopkins University Press, Baltimore, MD, USA, 2013), p. 73, <https://books.google.de/books?id=X5YfsuCWpxMC>.
- [47] K. Fan and A. J. Hoffman, Some metric inequalities in the space of matrices, *Proc. Am. Math. Soc.* **6**, 111 (1955).
- [48] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, Sample-optimal tomography of quantum states, *IEEE Trans. Inf. Theory* **63**, 5628 (2017).
- [49] R. O'Donnell and J. Wright, in *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, Cambridge, MA, USA, 2016), p. 899.
- [50] H. Yuen, An improved sample complexity lower bound for (fidelity) quantum state tomography, *Quantum* **7**, 890 (2023).
- [51] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, UK, 2013).
- [52] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (ACM, Phoenix, AZ, USA, 2019).