

Circuit cutting with classical side information

Christophe Piveteau^{1,*}, Lukas Schmitt^{1,2,*} and David Sutter^{1b,2}

¹*Institute for Theoretical Physics, ETH Zurich, 8093 Zürich, Switzerland*

²*IBM Quantum, IBM Research Europe–Zurich*



(Received 10 April 2025; accepted 25 June 2025; published 15 July 2025)

Circuit cutting is a technique for simulating large quantum circuits by partitioning them into smaller subcircuits, which can be executed on smaller quantum devices. The results from these subcircuits are then combined in classical postprocessing to accurately reconstruct the expectation value of the original circuit. Circuit cutting introduces a sampling overhead that grows exponentially with the number of gates and qubit wires that are cut. Many recently developed quasiprobabilistic circuit cutting techniques leverage classical side information, obtained from intermediate measurements within the subcircuits, to enhance the postprocessing step. In this work, we provide a formalization of general circuit cutting techniques utilizing side information through quantum instruments. With this framework, we analyze the advantage that classical side information provides in reducing the sampling overhead of circuit cutting. Surprisingly, we find that in certain scenarios, side information does not yield any reduction in sampling overhead, whereas in others it is essential for circuit cutting to be feasible at all. Furthermore, we present a lower bound for the optimal sampling overhead with side information that can be evaluated efficiently via semidefinite programming and improves on all previously known lower bounds.

DOI: [10.1103/38mx-36k6](https://doi.org/10.1103/38mx-36k6)

I. INTRODUCTION

A major difficulty in the realization of useful quantum computation is the limited number of available qubits. This problem is especially pressing, considering that quantum error correction is likely to require a large number of physical qubits to realize even a few logical qubits. One approach to solving this issue is to connect multiple quantum devices using quantum communication links over a network. However, such hardware is likely to remain out of reach in the coming years.

To address this shortcoming, recent years have seen the advent of circuit cutting techniques [1–8]. The idea is to partition large quantum circuits into smaller subcircuits that can each be executed on a smaller quantum device. Using appropriate classical postprocessing, the expectation value of an observable in the original circuit can be reconstructed with arbitrary precision—at the cost of a larger variance. This implies that to achieve a given (fixed) accuracy we require more executions of the circuit. This *sampling overhead* scales exponentially in the number of gates and qubit wires that were “cut” in the circuit partitioning. Despite its inevitable exponential overhead, circuit cutting is of high practical importance, and hence it is of interest to minimize the overhead as much as possible.

In this paper, we focus on the most popular and widely studied form of circuit cutting, which is based on the quasiprobability simulation technique. We note that other approaches do exist [1,9], though in comparison, they typically

lack the wide applicability. The key idea in quasiprobabilistic circuit cutting is to decompose the entangling gates across the circuit partitions into operations that act locally on each circuit partition. More mathematically, for a bipartite unitary channel $\mathcal{U}_{AB}$, one studies *quasiprobability decompositions* (QPD)

$$\mathcal{U}_{AB} = \sum_{i=1}^m a_i \mathcal{F}_i, \quad (1)$$

where $m \geq 1$ and $a \in \mathbb{R}^m$. Depending on whether we allow the circuit partitions to exchange classical communication or not, $\mathcal{F}_i$ are either local operations (LO) or local operations with classical communication (LOCC). Given such a QPD, the quasiprobability simulation technique can simulate the nonlocal unitary $\mathcal{U}$ by randomly replacing it by $\mathcal{F}_i$ and performing appropriate classical postprocessing of the circuit output. This leads to a sampling overhead of $\|a\|_1^2$ where $\|a\|_1 := \sum_i |a_i|$. When the QPDs of multiple nonlocal gates in the circuit are combined together, the ℓ_1 -norm of their coefficients scales multiplicatively, which leads to the exponential scaling of the sampling overhead in circuit cutting. For example, a circuit with n nonlocal gates $\mathcal{U}$ to be cut will incur a sampling overhead of $\|a\|_1^{2^n}$.

Clearly, it is desirable to pick a QPD that leads to the smallest possible sampling overhead. This optimal overhead is called the *quasiprobability extent*¹ $\gamma(\mathcal{E})$ of a bipartite channel $\mathcal{E}$, which is defined as the smallest ℓ_1 -norm $\|a\|_1$ over all valid QPDs of $\mathcal{E}$. To differentiate between situations in which the $\mathcal{F}_i$ are allowed to use classical communication or not, we denote the LO-assisted and LOCC-assisted extent by γ_{LO} and γ_{LOCC} , respectively. Clearly, as $\text{LO} \subseteq \text{LOCC}$, one has $\gamma_{\text{LOCC}} \leq \gamma_{\text{LO}}$.

*These authors contributed equally to this work.

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

¹Previous work sometimes called this quantity “ γ factor” [5,6,10]. We instead adopt the naming from Ref. [11].

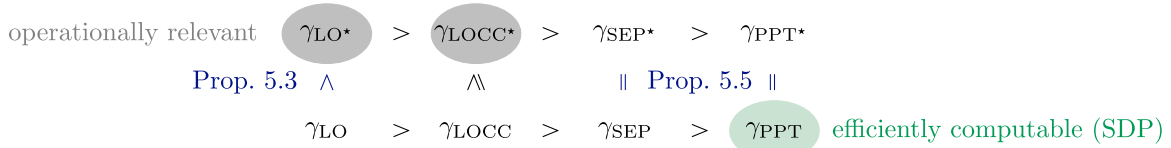


FIG. 1. Relations between different sampling overheads for CPTP maps. A strict inequality $>$ implies that there exists a CPTP map for which both extents differ. Here SEP denotes the set of all separable quantum channels. The sets LO^* , $LOCC^*$, SEP^* , and PPT^* are extensions of LO, LOCC, SEP, and PPT where we additionally allow for classical side information to be incorporated in the classical postprocessing. We refer to Sec. II for a precise definition of these sets. In Appendix A we show that all horizontal inequalities are strict.

The quasiprobability simulation method can be generalized by allowing the operations $\mathcal{F}_i$ to additionally produce classical information, e.g., by performing intermediate measurements, which is then incorporated into the classical postprocessing procedure. We refer to this improved technique as quasiprobability simulation with *classical side information*. The utilization of classical side information can be mathematically captured by allowing certain noncompletely positive and nontrace-preserving maps to be used in the QPD.

Example 1. Using classical side information, the nonpositive map $\mathcal{E} = \mathcal{E}^+ - \mathcal{E}^-$ can be realized if $\mathcal{E}^\pm$ are completely positive and sum up to a quantum channel. Consider the map

$$\rho_A \mapsto \mathcal{E}^+(\rho)_A |0\rangle\langle 0|_E + \mathcal{E}^-(\rho)_A \otimes |1\rangle\langle 1|_E, \quad (2)$$

then measuring the ancilar qubit E in the computational basis and weighting the final measurement outcome of the circuit by ± 1 realizes $\mathcal{E}$.

These enlarged sets of allowed operations will be denoted by LO^* and $LOCC^*$, and the corresponding extents by γ_{LO^*} and γ_{LOCC^*} , respectively. Since $LO \subseteq LO^*$ and $LOCC \subseteq LOCC^*$ we have $\gamma_{LO} \geq \gamma_{LO^*}$ and $\gamma_{LOCC} \geq \gamma_{LOCC^*}$. Given that there is not much reason not to allow for classical side information in practice, and we want to minimize the sampling overhead, γ_{LO^*} and γ_{LOCC^*} are the operationally relevant quantities for practical applications.

While specific quasiprobabilistic protocols utilizing classical side information have been around in literature for a while (see for example Ref. [12] in the context of quantum error mitigation and Refs. [2,3] for quasiprobabilistic circuit cutting), only recent work [5,6] formalized the optimal sampling overheads with side information through the quantities γ_{LO^*} and γ_{LOCC^*} .² Later, this formalization has been rephrased in different terms in Ref. [7]. However, the importance of classical side information for the purpose of circuit cutting has remained poorly understood. For instance, previous work does not address whether classical side information is necessary in the first place and whether it enables a smaller sampling overhead.

The quantities γ_{LO^*} and γ_{LOCC^*} have proven to be challenging to compute. This is not surprising, as they are closely related to entanglement measures [5], which are known to be NP-hard to evaluate in general [13]. We know the exact values of these quantities only for a few special cases, such as for unitaries with a Schmidt-like decomposition [6,7] and for pure states [5,14]. For more general quantum channels,

only little is known about their extent. Recent work [15] has made progress in this direction, by finding lower bounds of γ_{LOCC} through a relaxation of LOCC to the set of positive partial transpose channels (PPT), as $\gamma_{LOCC} \geq \gamma_{PPT}$ where γ_{PPT} is defined analogously to γ_{LO} and γ_{LOCC} . The quantity γ_{PPT} is expressible in terms of a semidefinite program (SDP) and can thus be efficiently evaluated numerically. Unfortunately, these PPT-based lower bounds are of limited applicability, as they only bound γ_{LOCC} , but not γ_{LOCC^*} , which is the physically relevant quantity. In other words, they do not make any statement about the performance of circuit cutting assisted by classical side information.

In this paper, we advance the understanding of circuit cutting with classical side information with following contributions:

(1) We mathematically formalize quasiprobabilistic circuit cutting with classical side information using the notion of *quantum instruments*. Essentially, we replace the channels $\mathcal{F}_i$ with quantum instruments, which describe the most general quantum processes that output classical information.³

(2) We show that classical side information is crucial for circuit cutting in the absence of classical communication. We prove that $\gamma_{LO}(\mathcal{E})$ is finite if and only if $\mathcal{E}$ is nonsignaling, whereas $\gamma_{LO^*}(\mathcal{E})$ is finite for all CPTP maps $\mathcal{E}$. We refer to Proposition 2 for the precise statement.

(3) We show that classical side information is useless in the PPT setting, i.e., $\gamma_{PPT}(\mathcal{E}) = \gamma_{PPT^*}(\mathcal{E})$ for all CPTP maps $\mathcal{E}$ (see Proposition 4). This implies that the SDP lower bound introduced in Ref. [15] is also applicable to γ_{LOCC^*} and not just γ_{LOCC} as discussed in Sec. VI.

(4) We study the PPT-based lower bound and show that it is tight in certain situations and always better than previously known lower bounds. We refer to Sec. VI and in particular to Corollary 2 for the rigorous statements.

The second and third result are summarized in Fig. 1.

The gray quantities γ_{LO^*} and γ_{LOCC^*} in Fig. 1 are the physically relevant quantities that capture the optimal overhead of circuit cutting without and with classical communication. The green quantity γ_{PPT} in Fig. 1 can be computed efficiently via a SDP and is a lower bound to the physically relevant quantities. While using $LOCC^*$ can in theory lead to a lower sampling cost than LO^* , the LO^* setting is still of practical relevance. Since it does not require classical communication, a single device can be used consecutively to simulate a larger one.

³Our work generalizes the formalism for classical side information in the quasiprobability simulation of nonphysical operations, which was recently introduced in Ref. [16].

²In Refs. [5,6] γ_{LO^*} and γ_{LOCC^*} were denoted by γ_{LO} and γ_{LOCC} .

II. PRELIMINARIES

A. Notation

For two finite-dimensional complex Hilbert spaces I, O , we denote by $L(I \rightarrow O)$ the set of linear operators from I to O and by $T(I \rightarrow O) := L(L(I \rightarrow I), L(O \rightarrow O))$ the set of linear superoperators from I to O . We denote the set of completely positive maps by $CP(I \rightarrow O)$ and the set of trace-preserving maps by $TP(I \rightarrow O)$. The set of quantum channels is $CPTP(I \rightarrow O) := CP(I \rightarrow O) \cap TP(I \rightarrow O)$. The set of completely positive trace nonincreasing maps is denoted by $CPTN(I \rightarrow O)$ and the set of Hermitian preserving superoperators by $HP(I \rightarrow O)$. The set of trace-preserving, Hermitian preserving maps is denoted $HPTP(I \rightarrow O)$. When the input and output systems are the same, we drop the arrow in the notation, for example $L(I \rightarrow I) = L(I)$ and $T(I \rightarrow I) = T(I)$. The set of density matrices on O is denoted by $D(O)$. For pure (rank one) states we use the symbol $\bar{D}(O)$. The Choi isomorphism $\mathcal{J} : T(I \rightarrow O) \rightarrow L(I' \otimes O)$ is defined as $\mathcal{J}(\mathcal{E}) := (\mathcal{I}_{I'} \otimes \mathcal{E})(|\Psi_{I'I}\rangle\langle\Psi_{I'I}|)$ where I' is a copy of I and $|\Psi_{I'I}\rangle := (\dim(I))^{-1/2} \sum_i |i\rangle_{I'} |i\rangle_I$ is a maximally entangled state across I' and I for some fixed orthonormal basis $|i\rangle$. It is well known that for a linear superoperator $\mathcal{E} \in T(I \rightarrow O)$ one has $\mathcal{E} \in CP(I \rightarrow O) \Leftrightarrow \mathcal{J}(\mathcal{E}) \geq 0$ and $\mathcal{E} \in TP(I \rightarrow O) \Leftrightarrow \text{tr}_O[\mathcal{J}(\mathcal{E})] = \frac{1}{\dim(I)} \mathbb{1}_{I'}$.

The most general measurement process in quantum mechanics can be described in terms of a channel $\mathcal{E}_{I \rightarrow OX} \in CPTP(I \rightarrow OX)$ which maps every state to a classical-quantum state,

$$\mathcal{E}_{I \rightarrow OX}(\rho_I) = \sum_i \mathcal{E}_i(\rho_I) \otimes |i\rangle\langle i|_X, \quad (3)$$

for some $\mathcal{E}_i \in T(I \rightarrow O)$, where I and O are the input and output quantum systems and X is a classical register that stores the measurement outcome. It is easy to check that all $\mathcal{E}_i$ must lie in $CPTN(I \rightarrow O)$ and, additionally, $\sum_i \mathcal{E}_i = \text{tr}_X \circ \mathcal{E}_{I \rightarrow OX} \in CPTP(I \rightarrow O)$. In contrast, all sets of operations $\mathcal{E}_i$ that meet these two properties describe a quantum measurement by Eq. (3). This motivates the definition of a (discrete) quantum instrument from I to O as a discrete family $(\mathcal{E}_i)_i$ of nonzero completely positive operators $\mathcal{E}_i \in CP(I \rightarrow O)$ such that $\sum_i \mathcal{E}_i \in CPTP(I \rightarrow O)$. The set of discrete quantum instruments from I to O is denoted $QI(I \rightarrow O)$.

We say a $(\mathcal{F}_j)_{j \in J} \in QI(I \rightarrow O)$ is a coarse-graining of $(\mathcal{E}_i)_{i \in K} \in QI(I \rightarrow O)$ if there exists a surjective map $f : K \rightarrow J$ such that $\forall j \in J : \mathcal{F}_j = \sum_{i: f(i)=j} \mathcal{E}_i$. The intuitive meaning is that $(\mathcal{F}_j)_j$ can be understood as the measurement $(\mathcal{E}_i)_i$ where some of the measurement outcomes are treated as the same. So, the instrument $(\mathcal{F}_j)_j$ can be realized by first implementing $(\mathcal{E}_i)_i$ and then discarding some classical information.

A bipartite state $\rho \in D(AB)$ is not entangled (also called separable) if it can be written as

$$\rho = \sum_i p_i \sigma_i \otimes \tau_i, \quad (4)$$

for some probability distribution $(p_i)_i$ and $\sigma_i \in D(A)$, $\tau_i \in D(B)$. The set of separable states is denoted by $SEP(A, B)$. Testing for separability is challenging, as it is not a convex constraint. For this reason, a common technique is to relax

this constraint by defining the larger set

$$\text{PPT}(A, B) := \{\rho \in D(AB) : \rho^{T_B} \geq 0\}, \quad (5)$$

where ρ^{T_B} denotes the partial transposition of ρ on the system B . Indeed, $SEP(A, B) \subset \text{PPT}(A, B)$ as any separable state has a non-negative partial transpose.⁴ This is the well-known PPT criterion which states that any state with positive partial transpose must be entangled [17,18].

B. Quasiprobability simulation

Suppose we are given a quantum circuit that prepares some state and then measures an observable. We are interested in predicting the expectation value of this observable with a quantum computer that is restricted in the sense that it cannot realize a full universal gate set. Let us separate the set of all quantum operations into the “free” operations (denoted by the set S) which the quantum computer can physically execute, as well as “nonfree” operations.⁵ If the desired quantum circuits contains nonfree operations, then it is *a priori* impossible to directly execute the circuit on the given quantum computer. However, using the quasiprobability simulation technique, it is still possible to evaluate the expectation value of the original circuit with the restricted quantum computer, albeit with a sampling overhead.

Quasiprobability simulation has been applied to a variety of different applications, each corresponding to a different choice of the set of free operations S . For example, it can be used to simulate near-Clifford circuits on a classical computer [19,20] (here S is chosen to be Clifford operations) and for quantum error mitigation [12,21] (here S is chosen to be the set of noisy operations executable by the quantum computer). In the context of quasiprobabilistic circuit cutting, the set S is chosen to contain local operations across the circuit bipartitions, and quasiprobability simulation allows for the simulation of nonlocal gates [3–5]. We will elaborate more on the application to circuit cutting in Sec. IV, and in the following, we briefly summarize how the quasiprobability technique operates. For a comprehensive treatment, we refer the reader to Ref. [11]. Suppose that our task is to simulate some nonfree operation $\mathcal{E} \in T(I \rightarrow O)$ using free operations. Furthermore, assume that we can decompose $\mathcal{E}$ as a linear combination using elements $\mathcal{F}_i \in S \subset T(I \rightarrow O)$,

$$\mathcal{E} = \sum_{i=1}^m a_i \mathcal{F}_i, \quad (6)$$

where $a_i \in \mathbb{R}$. This decomposition is called a QPD, because some of the coefficients a_i might generally be negative. Using some simple algebra, we can rewrite this QPD as to remove

⁴Later, we will also denote the set of separable and PPT channels with the symbols SEP and PPT . It will always be clear from context which set is considered.

⁵The expression “free” is inspired by the literature of quantum resource theories. One can often equate the decomposition set S with the free operations of some quantum resource theory; for more details, see Ref. [11].

the negativity form the weight and make it a proper probabilistic mixture,

$$\mathcal{E} = \sum_{i=1}^m p_i \mathcal{F}_i \text{sign}(a_i) \|a\|_1, \quad (7)$$

with the probability distribution $p_i = |a_i|/\|a\|_1$. The idea of quasiprobability simulation is to probabilistically replace the operation $\mathcal{E}$ by $\mathcal{F}_i$ with probability p_i , and then weight the final measurement outcome by $\text{sign}(a_i)\|a\|_1$. To see that this indeed provides us with an unbiased estimate of the expectation value, let us consider the simplest instance where the circuit consists purely of the operation $\mathcal{E}$ followed by the measurement of the observable O . Then,

$$\begin{aligned} \text{tr}[O\mathcal{E}(\rho)] &= \sum_{i=1}^m a_i \text{tr}[O\mathcal{F}_i(\rho)] \\ &= \sum_{i=1}^m p_i \text{tr}[O\mathcal{F}_i(\rho)] \text{sgn}(a_i) \|a\|_1. \end{aligned} \quad (8)$$

Hence, the random replacement of $\mathcal{E}$ by $\mathcal{F}_i$, coupled with the appropriate postprocessing, provides an unbiased estimator for $\text{tr}[O\mathcal{E}(\rho)]$. However, since we are scaling up the circuit output by the factor $\pm\|a\|_1$, this estimator generally has a higher variance compared to directly executing the target circuit in the first place. Therefore, a higher number shots is needed in order to reach the same confidence for the expectation value.

Proposition 1. Let $\epsilon > 0$ and $\delta \in (0, 1]$. Using $\frac{2\|a\|_1^2}{\epsilon^2} \log(\frac{2}{\delta})$ shots, quasiprobability simulation produces an estimate for $\text{tr}[O\mathcal{E}(\rho)]$ up to an additive error ϵ with probability at least $1 - \delta$.

This proposition is a simple consequence of Hoeffding's inequality, and we refer the reader to Refs. [12,19,21] for more details. Since the number of shots required increases multiplicatively in $\|a\|_1^2$, it is desirable to pick a QPD that has the smallest possible ℓ_1 -norm of its coefficients. Given a fixed set S of operations that our computer can perform, we define the extent of a map $\mathcal{E}$ as the smallest such achievable ℓ_1 -norm.

Definition 1. For a decomposition set $S \subset \mathcal{T}(I \rightarrow O)$ and a target operation $\mathcal{E} \in \mathcal{T}(I \rightarrow O)$, we define the *quasiprobability extent* of $\mathcal{E}$ with respect to S as

$$\gamma_S(\mathcal{E}) := \inf \left\{ \sum_{i=1}^m |a_i| : \mathcal{E} = \sum_{i=1}^m a_i \mathcal{F}_i \text{ where } m \geq 1, \mathcal{F}_i \in S \text{ and } a_i \in \mathbb{R} \right\}. \quad (9)$$

If no valid decomposition exists, we define the extent to be infinite. An infinite extent indicates that a quasiprobabilistic simulation of $\mathcal{E}$ with a decomposition set S is not possible.

Previously, we have illustrated how quasiprobability simulation can estimate the expectation value of an observable on a circuit which contains one single nonfree operation $\mathcal{E}$. The technique can naturally be applied to circuits which contain multiple such gates, given that each gate exhibits its own QPD into operations in S . The main idea is to simply combine the individual QPDs into a large QPD of the total circuit, such that the previous consideration is again applicable. As an example,

suppose we have three operations $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$ that do not lie in S and we want to implement the circuit

$$\mathcal{U}_4 \circ \mathcal{E}_3 \circ \mathcal{U}_3 \circ \mathcal{E}_2 \circ \mathcal{U}_2 \circ \mathcal{E}_1 \circ \mathcal{U}_1, \quad (10)$$

where $\mathcal{U}_i$ denote concatenations of operations that lie in S . Assuming we have the QPDs $\mathcal{E}_j = \sum_i a_i^{(j)} \mathcal{F}_i^{(j)}$, Eq. (10) can be rewritten as

$$\begin{aligned} &\sum_{i_1, i_2, i_3} p_{i_1}^{(1)} p_{i_2}^{(2)} p_{i_3}^{(3)} \mathcal{U}_4 \circ \mathcal{F}_{i_3}^{(3)} \circ \mathcal{U}_4 \circ \mathcal{F}_{i_2}^{(2)} \circ \mathcal{U}_2 \circ \mathcal{F}_{i_1}^{(1)} \\ &\circ \mathcal{U}_1 \text{sgn}(a_{i_1}^{(1)} a_{i_2}^{(2)} a_{i_3}^{(3)}) \|a^{(1)}\|_1 \|a^{(2)}\|_1 \|a^{(3)}\|_1, \end{aligned} \quad (11)$$

where $p_i^{(j)} := |a_i^{(j)}|/\|a^{(j)}\|_1$. The Monte Carlo sampling procedure in this case works as follows: For every instance of an unsupported operation $\mathcal{E}_j$ in the circuit, we separately and independently sample a random gate $\mathcal{F}_i^{(j)}$ according to $p_i^{(j)}$ to replace it. The sampled indices (i_1, i_2 , and i_3) are stored, and at the end of the circuits we weight the measurement outcome by $\pm\|a^{(1)}\|_1 \|a^{(2)}\|_1 \|a^{(3)}\|_1$. Note how the ℓ_1 -norms of the individual decompositions combine in a multiplicative fashion, so the overall sampling overhead generally scales exponentially in the number of unsupported operations.

The possibly infinite number of elements that a QPD can contain might seem daunting to deal with. Fortunately, in many cases the decomposition set S is convex, allowing for a simpler characterization with QPDs containing only two elements.

Lemma 1. Let $S \subset \mathcal{T}(I \rightarrow O)$ be a convex set and $\mathcal{E} \in \mathcal{T}(I \rightarrow O)$. Then

$$\begin{aligned} \gamma_S(\mathcal{E}) &= \inf \{a^+ + a^- : \mathcal{E} = a^+ \mathcal{E}^+ - a^- \mathcal{E}^- \text{ where } \\ &a^\pm \geq 0 \text{ and } \mathcal{E}^\pm \in S\}. \end{aligned} \quad (12)$$

Proof. It suffices to show that any QPD of $\mathcal{E}$ with respect to S induces another QPD that has at most two elements and the same ℓ_1 -norm of its coefficients. Let $\mathcal{E} = \sum_i a_i \mathcal{F}_i$ be one such QPD. Define

$$\begin{aligned} a^+ &:= \sum_{i: a_i \geq 0} a_i, \quad a^- := - \sum_{i: a_i < 0} a_i, \\ \mathcal{E}^+ &:= \frac{1}{a^+} \sum_{i: a_i \geq 0} a_i \mathcal{F}_i, \quad \text{and} \quad \mathcal{E}^- := \frac{1}{a^-} \sum_{i: a_i < 0} a_i \mathcal{F}_i. \end{aligned} \quad (13)$$

Clearly, $\mathcal{E} = a^+ \mathcal{E}^+ - a^- \mathcal{E}^-$ and $a^+ + a^- = \sum_i |a_i|$ and $\mathcal{E}^\pm \in S$ by convexity. ■

The quasiprobability simulation technique cannot only simulate nonfree quantum channels but also nonfree states. Indeed, states can be interpreted as a special case of quantum channels with trivial input spaces, so the above methods can be readily applied.⁶ Correspondingly, the quasiprobability extent of a state $\rho \in \mathcal{D}(A) \equiv \text{CPTP}(\mathbb{C} \rightarrow A)$ with respect to

⁶In some applications of quasiprobability simulation, the focus more commonly lies on the simulation of nonfree states (instead of channels). For instance, the classical simulation of Clifford+T circuits can be achieved by simulating magic states with stabilizer states [20,22].

some decomposition set $S \subset T(\mathbb{C} \rightarrow A)$ is given by

$$\gamma_S(\rho) := \inf \left\{ \sum_{i=1}^m |a_i| : \rho = \sum_{i=1}^m a_i \sigma_i \text{ where } m \geq 1, \sigma_i \in S \text{ and } a_i \in \mathbb{R} \right\}. \quad (14)$$

If the decomposition set S is convex, one can connect the extent to a corresponding robustness measure. The robustness of a state ρ with respect to a decomposition set S is given by

$$R_S(\rho) := \min_{\sigma \in S, t \geq 0} \left\{ t : \frac{\rho + t\sigma}{1+t} \in S \right\}. \quad (15)$$

Lemma 2. Let S be a convex set and ρ a quantum state. Then,

$$\gamma_S(\rho) = 1 + 2R_S(\rho). \quad (16)$$

Proof. Based on Lemma 1, it is easy to see that any optimal decomposition of one quantity induces a valid decomposition for the other. We refer to Ref. [11, Lemma 3.17] for a more detailed proof that also holds for channels. ■

III. QUASIPROBABILITY SIMULATION WITH CLASSICAL SIDE INFORMATION

It might seem natural to restrict ourselves to considering decomposition sets S that are subsets of the completely positive trace-preserving (CPTP) maps, since these are precisely the physically realizable operations. In this section, we will see that it is useful to allow S to also contain certain trace nonincreasing maps and even noncompletely positive maps. This should be understood as a mathematical trick that allows us to capture the notion of *classical side information* in the quasiprobability simulation procedure. As explained in Sec. II B, the procedure operates by randomly replacing nonfree operations by free ones and then appropriately multiplying the output of the circuit by some real scalar. This technique can be augmented by allowing this postprocessing step to additionally multiply the output by some scalar which may depend on midcircuits measurements. Let us first consider a few explanatory examples before presenting the trick in full generality.

As an example, consider that we want to simulate the one-qubit trace nonincreasing operation $\Pi_0(\cdot) = |0\rangle\langle 0|(\cdot)|0\rangle\langle 0|$ which maps $\rho \mapsto \langle 0|\rho|0\rangle|0\rangle\langle 0|$. This map cannot be written as a linear combination of CPTP maps, because any such linear combination would be proportional to a trace-preserving map. Hence, Π_0 cannot be simulated if S only contains CPTP maps (i.e., if no classical side information is utilized). Yet still, it can be simulated by performing a measurement in the computational basis described by

$$\mathcal{M}(\rho) := \sum_{i=0}^1 |i\rangle\langle i|\rho|i\rangle\langle i| \otimes |i\rangle\langle i|_X \quad (17)$$

which stores the outcome in a classical register X . Notice that one can write $\Pi_0(\rho) = (\mathbb{1} \otimes \langle 0|_X) \mathcal{M}(\rho) (\mathbb{1} \otimes |0\rangle_X)$. If we now insert this in the term for the expectation value of some observable O on some circuit with initial state ρ_{in} and operations $\mathcal{U}_2 \circ \Pi_0 \circ \mathcal{U}_1$ (here $\mathcal{U}_i$ stand for all other operations

that occur before and after in the circuit), then the $\langle 0|_X$ and $|0\rangle_X$ can be absorbed into the observable, i.e.,

$$\text{tr}[O \cdot \mathcal{U}_2 \circ \Pi_0 \circ \mathcal{U}_1(\rho)] = \text{tr}[(O \otimes |0\rangle\langle 0|_X) \mathcal{U}_2 \circ \mathcal{M} \circ \mathcal{U}_1(\rho)]. \quad (18)$$

Hence, the expectation value can be estimated as follows: We replace the Π_0 operation in the circuit by the computational basis measurement $\mathcal{M}$. For every circuit execution, we weight the measurement output either by 1 or 0, depending on the midcircuit measurement outcome. Note that this is not post-selection, as the “discarded” runs are still taken into account when averaging over multiple runs, they just have value zero. Equation (18) guarantees that this procedure constitutes an unbiased estimator.

It is illustrative to consider a second example, this time of the noncompletely positive map $\mathcal{E} = \Pi_0 - \Pi_1$ where we analogously define $\Pi_1(\cdot) = |1\rangle\langle 1|(\cdot)|1\rangle\langle 1|$. Here again, if we only care about the final observable, then we can simulate this $\mathcal{E}$ by replacing it with a computational basis measurement,

$$\begin{aligned} \text{tr}[O \cdot \mathcal{U}_2 \circ (\Pi_0 - \Pi_1) \circ \mathcal{U}_1(\rho)] \\ = \text{tr}[(O \otimes (|0\rangle\langle 0|_X - |1\rangle\langle 1|_X)) \cdot \mathcal{U}_2 \circ \mathcal{M} \circ \mathcal{U}_1(\rho)] \end{aligned} \quad (19)$$

and by weighting the final circuit observable outcome by either +1 or −1 depending on the intermediate measurement outcome.

The main insights of the previous two examples is that *for the purpose of estimating some expectation value* one can use classical side information to simulate non-CPTP maps by weighting the final observable measurement by some factor depending on the intermediate measurement outcome. The most general operation that produces such classical information is described by a quantum instrument. Given a discrete quantum instrument $(\mathcal{G}_j)_j$ that our computer can physically perform, this technique allows us to effectively run any operation $\sum_j b_j \mathcal{G}_j$ for some weighting factors $b_j \in \mathbb{R}$. This is useful for quasiprobability simulation, because when we now decompose our target operation $\mathcal{E} = \sum_i a_i \mathcal{F}_i$, the $\mathcal{F}_i$ do not need to be CPTP, but instead they could rather be of this more general form. This might possibly allow us to achieve a smaller extent. Note that in order for Proposition 1 to still hold, we cannot weight the outcome by a factor b_j larger than 1 in absolute value, as that would increase the variance of the estimators. For this reason, we have to restrict ourselves coefficients fulfilling $|b_j| \leq 1$.

In summary, when characterizing the capabilities of a restricted quantum computer to perform quasiprobability simulation with classical side information, we should not consider the set of free CPTP maps that the computer can perform. Instead, it is more insightful to consider the set of free quantum instruments that the computer can realize. Given a quantum computer that can realize some set of quantum instruments $\mathcal{Q}$, the associated decomposition set (which captures the capability of using side information) is given as follows.

Definition 2. Let $Q \subset \text{QI}(I \rightarrow O)$ be a set of discrete quantum instruments. We define the set

$$\mathcal{L}[Q] := \left\{ \sum_i a_i \mathcal{F}_i : (\mathcal{F}_i)_i \in Q, a_i \in [-1, 1] \right\}. \quad (20)$$

By our above discussion, the optimal simulation overhead (with classical side information) for an operation $\mathcal{E}$ is thus characterized by $\gamma_{\mathcal{L}[Q]}(\mathcal{E})$. For the rest of this section, we provide some mathematical properties of $\mathcal{L}[Q]$.

Definition 3. Let $Q \subset \text{QI}(I \rightarrow O)$ be a set of quantum instruments. We say that

(1) Q is *coarse-grainable* if $\forall (\mathcal{E}_i)_i \in Q$, any coarse-graining of $(\mathcal{E}_i)_i$ is also in Q .

(2) Q is *trivially fine-grainable* if $\forall (\mathcal{E}_i)_i \in Q$ we have $((1-p_i)\mathcal{E}_i)_i \cup (p_i\mathcal{E}_i)_i \in Q$ for any choice of probabilities $p_i \in [0, 1]$.

(3) Q is *closed under mixture* if for any two instruments $(\mathcal{E}_i)_i$ and $(\mathcal{F}_j)_j$ in Q , their convex mixture $(p\mathcal{E}_i)_i \cup ((1-p)\mathcal{F}_j)_j$ is also in Q , for any $p \in [0, 1]$.

It is quite natural for the set of free quantum instruments to fulfill these three properties, and they hold in most physically relevant settings. Q being coarse-grainable means that one can discard classical information and treat different classical outcomes as the same. Trivial finegrainability essentially means that one can sample a random bit with the probability depending on the observed measurement outcome. Similarly, when Q is closed under mixture, then randomly choosing between two free quantum instruments is itself also a free quantum instrument in Q .

At first glance, using the decomposition set $\mathcal{L}[Q]$ seems to add a lot of complexity, as now every element of the QPD could consist of a possibly infinite sum of quantum instrument elements. However, this turns out not to be the case. Indeed, it is generally sufficient to consider two-element instruments and coefficients $a_i = \pm 1$ when the set Q fulfills the aforementioned natural assumptions.

Lemma 3. If $Q \subset \text{QI}(A)$ is coarse-grainable and trivially fine-grainable, then

$$\mathcal{L}[Q] = \{\mathcal{E}^+ - \mathcal{E}^- : (\mathcal{E}^+, \mathcal{E}^-) \in Q\}. \quad (21)$$

Proof. Consider an element $\mathcal{E} = \sum_i a_i \mathcal{E}_i \in \mathcal{L}[Q]$ where $(\mathcal{E}_i)_i \in Q$. By the trivially fine-grainable property of Q , we can also express $\mathcal{E}$ as a weighted sum of another instrument in Q , but only with coefficients $+1$, -1 , and 0 :

$$\mathcal{E} = \sum_i \text{sgn}(a_i)(|a_i|\mathcal{E}_i) + \sum_i 0 \cdot ((1-|a_i|)\mathcal{E}_i) \quad (22)$$

We now coarse-grain this instrument into a three-element instrument $(\mathcal{E}^+, \mathcal{E}^-, \mathcal{E}^0)$,

$$\begin{aligned} \tilde{\mathcal{E}}^+ &:= \sum_{i:a_i \geq 0} |a_i| \mathcal{E}_i, & \tilde{\mathcal{E}}^- &:= \sum_{i:a_i < 0} |a_i| \mathcal{E}_i, \\ \mathcal{E}^0 &:= \sum_i (1 - |a_i|) \mathcal{E}_i, \end{aligned} \quad (23)$$

which is still in Q by the coarse-grainability property. We thus have

$$\mathcal{E} = \tilde{\mathcal{E}}^+ - \tilde{\mathcal{E}}^- + 0 \cdot \mathcal{E}^0. \quad (24)$$

Finally, the fine-grainability and coarse-grainability properties imply $(\tilde{\mathcal{E}}^+, \tilde{\mathcal{E}}^-, \frac{1}{2}\mathcal{E}^0, \frac{1}{2}\mathcal{E}^0)$ and thus $(\tilde{\mathcal{E}}^+ + \frac{1}{2}\mathcal{E}^0, \tilde{\mathcal{E}}^- + \frac{1}{2}\mathcal{E}^0)$ are quantum instruments in Q . The desired statement thus follows from

$$\mathcal{E} = (\tilde{\mathcal{E}}^+ + \frac{1}{2}\mathcal{E}^0) - (\tilde{\mathcal{E}}^- + \frac{1}{2}\mathcal{E}^0). \quad (25)$$

We also note that any set Q being closed under mixture provides useful structure for the induced decomposition set.

Lemma 4. Let $Q \subset \text{QI}(I \rightarrow O)$ be a set of quantum instruments that is closed under mixture. Then $\mathcal{L}[Q]$ is convex.

Proof. Let $\mathcal{E}$ and $\mathcal{F}$ be two elements in $\mathcal{L}[Q]$. By definition, we can write them as $\mathcal{E} = \sum_i b_i \mathcal{E}_i$ and $\mathcal{F} = \sum_j c_j \mathcal{F}_j$ for two quantum instruments $(\mathcal{E}_i)_i, (\mathcal{F}_j)_j$ in Q . Any convex mixture of the two elements is also in $\mathcal{L}[Q]$, because we can write

$$(1-p)\mathcal{E} + p\mathcal{F} = \sum_i b_i(1-p)\mathcal{E}_i + \sum_j c_j p\mathcal{F}_j, \quad (26)$$

where $p \in [0, 1]$ and $(p\mathcal{E}_i)_i \cup ((1-p)\mathcal{F}_j)_j$ is in Q due to it being closed under mixture.

In fact, if a set of quantum instruments Q is closed under mixture, then the induced set of quantum channels is also convex.

Lemma 5. Let $Q \subset \text{QI}(I \rightarrow O)$ be a set of quantum instruments and $S^* := \mathcal{L}[Q]$. We denote by $S := \{\sum_i \mathcal{E}_i | (\mathcal{E}_i)_i \in Q\}$ the set of quantum channels induced by Q . If S^* is convex, then so is S .

Proof. Consider two channels $\mathcal{E}, \mathcal{F} \in S$. We know that any convex mixture $(1-p)\mathcal{E} + p\mathcal{F}$ lies in S^* , and we now need to show that it also lies in S . By the definition of S^* , we can write

$$(1-p)\mathcal{E} + p\mathcal{F} = \sum_j b_j \mathcal{G}_j \quad (27)$$

for some $b_j \in [-1, 1]$ and $(\mathcal{G}_j)_j \in Q$. Taking the Choi isomorphism on both sides, as well as the partial trace tr_O thereof, we obtain

$$\frac{1}{\dim(I)} \mathbb{1} = \sum_j b_j \text{tr}_O[\mathcal{J}(\mathcal{G}_j)]. \quad (28)$$

At the same time, we know that $\sum_j \mathcal{G}_j \in \text{CPTP}(I \rightarrow O)$, so

$$\frac{1}{\dim(I)} \mathbb{1} = \sum_j b_j \text{tr}_O[\mathcal{J}(\mathcal{G}_j)] \leq \sum_j \text{tr}_O[\mathcal{J}(\mathcal{G}_j)] = \frac{1}{\dim(I)} \mathbb{1}. \quad (29)$$

This implies that for all j we have $b_j = 1$, unless $\mathcal{G}_j$ is itself zero. Therefore, $(1-p)\mathcal{E} + p\mathcal{F} \in S$. ■

IV. QUASIPROBABILISTIC CIRCUIT CUTTING

In this section, we elaborate how the quasiprobability technique can be used to simulate a large quantum computer with a small one. Suppose one wants to run a quantum circuit on a given quantum device, but its number of available qubits is too small to fit the circuit. The premise of *circuit cutting* is to subdivide the circuit into smaller partitions which are each small enough to fit on the available quantum computer. The goal is to reconstruct the outcome of the original circuit by executing the smaller partitions on a small quantum computer and performing appropriate classical postprocessing.

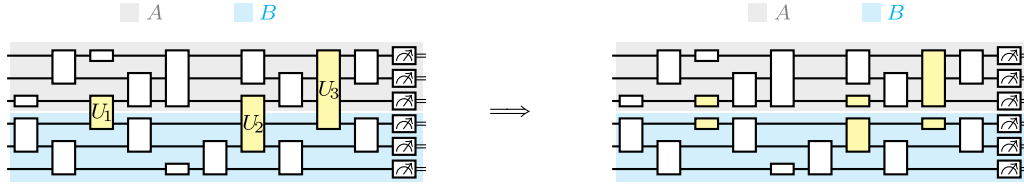


FIG. 2. The nonlocal circuit on the left can be simulated with local circuits on the right using quasiprobability simulation. Each nonlocal gate U_1 , U_2 , and U_3 is simulated using local operations.

It is natural to realize circuit cutting through quasiprobability simulation. Consider for instance the setup depicted in Fig. 2. The qubits in a circuit are grouped into two systems A and B . Every gate that acts nonlocally across A and B is considered a nonfree operation, and we will simulate them using operations that only act locally on each partition.

It is common to distinguish between two settings where classical communication between the subpartitions is either allowed or not allowed. Mathematically, the distinction between these two situations is captured by choosing the decomposition set S to be LO or LOCC, respectively, and LO^* and LOCC^* when classical side information is also allowed. A precise mathematical definition of these four sets will be given further below.

Classical communication could potentially lead to a lower simulation overhead for certain bipartite channels $\mathcal{E}$. In Appendix A, we show that there exist bipartite channels $\mathcal{E}$ such that $\gamma_{\text{LOCC}^*}(\mathcal{E}) < \gamma_{\text{LO}^*}(\mathcal{E})$. It is known that $\gamma_{\text{LOCC}^*}(\mathcal{U}) = \gamma_{\text{LO}^*}(\mathcal{U})$ for Clifford gates [5], all two-qubit unitary channels $\mathcal{U}$ and KAK-like generalizations thereof [6]. It is unknown if this equality is true for arbitrary unitaries.

The use of classical communication comes at an additional price in terms of practical implementation costs. Without classical communication, the individual subcircuits could be run

sequentially and in principle on the same quantum device. However, once classical communication is involved, multiple small quantum processors will generally be required to execute the subcircuits simultaneously.

The set LOCC is notoriously hard to work with due to its intricate mathematical structure. A common strategy to ease this difficulty is to replace LOCC with a relaxation of separable channels SEP or positive partial transpose channels PPT, which are strict supersets of LOCC with simpler characterization. A precise mathematical definition will follow later, but we already point out that

$$\text{LO} \subset \text{LOCC} \subset \text{SEP} \subset \text{PPT} \quad (30)$$

and

$$\text{LO}^* \subset \text{LOCC}^* \subset \text{SEP}^* \subset \text{PPT}^*, \quad (31)$$

which in turn implies the relations between the associated quasiprobability extents depicted in Fig. 3.⁷ We highlight that compared to LO and LOCC, SEP and PPT have no operational meaning for the purpose of circuit cutting and instead only serve as a mathematical tool to find lower bounds.

We now introduce the quantum instruments of these four settings in detail:

$$\text{QI}_{\text{LO}(AB \rightarrow A'B')} := \{(\mathcal{E}_i \otimes \mathcal{F}_j)_{i,j} : (\mathcal{E}_i)_i \in \text{QI}(A \rightarrow A'), (\mathcal{F}_j)_j \in \text{QI}(B \rightarrow B')\}, \quad (32)$$

$$\text{QI}_{\text{SEP}(AB \rightarrow A'B')} := \left\{ (\mathcal{E}_i)_i \in \text{QI}(AB \rightarrow A'B') : \forall i : \frac{\mathcal{J}(\mathcal{E}_i)}{\text{tr}[\mathcal{J}(\mathcal{E}_i)]} \in \text{SEP}(AA', BB') \right\}, \quad (33)$$

$$\text{QI}_{\text{PPT}(AB \rightarrow A'B')} := \{(\mathcal{E}_i)_i \in \text{QI}(AB \rightarrow A'B') : \forall i : \mathcal{J}(\mathcal{E}_i)^{T_B} \geq 0\}. \quad (34)$$

The precise definition of the LOCC instruments QI_{LOCC} is technical and not necessary for our purposes here. We refer the reader to Ref. [23] for a precise mathematical treatment of QI_{LOCC} . All these instruments induce a set of achievable channels, i.e.,

$$\text{LO}(AB \rightarrow A'B') := \left\{ \sum_i \mathcal{E}_i : (\mathcal{E}_i)_i \in \text{QI}_{\text{LO}(AB \rightarrow A'B')} \right\} = \{\mathcal{E} \otimes \mathcal{F} : \mathcal{E} \in \text{CPTP}(A \rightarrow A'), \mathcal{F} \in \text{CPTP}(B \rightarrow B')\} \quad (35)$$

$$\text{LOCC}(AB \rightarrow A'B') := \left\{ \sum_i \mathcal{E}_i : (\mathcal{E}_i)_i \in \text{QI}_{\text{LOCC}(AB \rightarrow A'B')} \right\}, \quad (36)$$

$$\text{SEP}(AB \rightarrow A'B') := \left\{ \sum_i \mathcal{E}_i : (\mathcal{E}_i)_i \in \text{QI}_{\text{SEP}(AB \rightarrow A'B')} \right\} = \{\mathcal{E} \in \text{CPTP}(AB \rightarrow A'B') : \mathcal{J}(\mathcal{E}) \in \text{SEP}(AA', BB')\}, \quad (37)$$

$$\text{PPT}(AB \rightarrow A'B') := \left\{ \sum_i \mathcal{E}_i : (\mathcal{E}_i)_i \in \text{QI}_{\text{PPT}(AB \rightarrow A'B')} \right\} = \{\mathcal{E} \in \text{CPTP}(AB \rightarrow A'B') : \mathcal{J}(\mathcal{E}) \in \text{PPT}(AA', BB')\}. \quad (38)$$

⁷When considering the extent of some bipartite channel $\mathcal{E} \in \text{CPTP}(AB \rightarrow A'B')$, we will typically drop the systems AB in the subscript [e.g., $\gamma_{\text{LOCC}}(\mathcal{E})$ instead of $\gamma_{\text{LOCC}(AB \rightarrow A'B')}(\mathcal{E})$] to simplify the notation. The implied systems will be clear from context.

$$\begin{array}{ccccccc}
\gamma_{\text{LO}}^* & \geq & \gamma_{\text{LOCC}}^* & \geq & \gamma_{\text{SEP}}^* & \geq & \gamma_{\text{PPT}}^* \\
\wedge & & \wedge & & \wedge & & \wedge \\
\gamma_{\text{LO}} & \geq & \gamma_{\text{LOCC}} & \geq & \gamma_{\text{SEP}} & \geq & \gamma_{\text{PPT}}
\end{array}$$

FIG. 3. Trivial relations of extents for different settings. The inequalities follow from the inclusion relations of the different decomposition sets [see Eqs. (30) and (31)]. A main result of this work is a more refined version of this diagram, which is depicted in Fig. 1.

Following the procedure from Sec. III, the corresponding decomposition sets capturing the notion of classical side information are given by

$$\text{LO}^*(AB \rightarrow A'B') := \mathcal{L}[\text{QI}_{\text{LO}(AB \rightarrow A'B')}], \quad (39)$$

$$\begin{aligned}
\text{LOCC}^*(AB \rightarrow A'B') &:= \mathcal{L}[\text{QI}_{\text{LOCC}(AB \rightarrow A'B')}] \\
&= \{\mathcal{E}^+ - \mathcal{E}^- : (\mathcal{E}^+, \mathcal{E}^-) \in \text{QI}_{\text{LOCC}(AB \rightarrow A'B')}\}, \quad (40)
\end{aligned}$$

$$\begin{aligned}
\text{SEP}^*(AB \rightarrow A'B') &:= \mathcal{L}[\text{QI}_{\text{SEP}(AB \rightarrow A'B')}] \\
&= \{\mathcal{E}^+ - \mathcal{E}^- : (\mathcal{E}^+, \mathcal{E}^-) \in \text{QI}_{\text{SEP}(AB \rightarrow A'B')}\}, \quad (41)
\end{aligned}$$

$$\begin{aligned}
\text{PPT}^*(AB \rightarrow A'B') &:= \mathcal{L}[\text{QI}_{\text{PPT}(AB \rightarrow A'B')}] \\
&= \{\mathcal{E}^+ - \mathcal{E}^- : (\mathcal{E}^+, \mathcal{E}^-) \in \text{QI}_{\text{PPT}(AB \rightarrow A'B')}\}, \quad (42)
\end{aligned}$$

where we used Lemma 3 for the characterization of LOCC^* , SEP^* , and PPT^* . Indeed, it follows immediately from the definition of QI_{SEP} and QI_{PPT} that they are coarse-grainable and trivially fine-grainable. For QI_{LOCC} , the coarse-grainability follows directly from its definition and trivial fine-grainability essentially can be seen by concatenating any LOCC protocol with one of the two parties choosing a random bit with probability based on all previously obtained measurement outcomes. Furthermore, LOCC^* , SEP^* , and PPT^* are convex by Lemma 4 because the corresponding sets of quantum instruments are closed under mixture.⁸ These properties do unfortunately not hold for QI_{LO} , making its characterization slightly more challenging—this will not be a problem for our purposes in this paper.

The four proposed settings constitute a hierarchy in the sense that

$$\text{QI}_{\text{LO}} \subset \text{QI}_{\text{LOCC}} \subset \text{QI}_{\text{SEP}} \subset \text{QI}_{\text{PPT}}, \quad (43)$$

which therefore implies Eqs. (30) and (31) and the relations in Fig. 3. One of the main technical results of this work is to refine these relations with those seen in Fig. 1.

We highlight that the two quantities with practical relevance for circuit cutting are γ_{LO}^* and γ_{LOCC}^* . The quantities γ_{LO} and γ_{LOCC} are not as natural, as there is little reason not

to use the power classical side information, especially in the LOCC setting where intermediate measurements are required in the LOCC protocol itself. We still study these two quantities to understand the improvement that classical side information can enable. The quantities γ_{PPT} , γ_{SEP} , γ_{PPT}^* , and γ_{SEP}^* have no physical meaning and will be used purely as a mathematical tool to provide lower bounds for the other quantities.

V. UTILITY OF CLASSICAL SIDE INFORMATION

In this section, we study whether the usage of classical side information allows for a smaller sampling overhead. Before we consider general channels, let us consider the simpler case of states. Here classical side information does not provide any utility, and in fact, we get the same extent in the LOCC, LO, and SEP settings.

Lemma 6. For any bipartite density matrix $\rho \in \text{D}(AB)$ we have

$$\begin{aligned}
\gamma_{\text{LO}}(\rho) &= \gamma_{\text{LO}}^*(\rho) = \gamma_{\text{LOCC}}(\rho) = \gamma_{\text{LOCC}}^*(\rho) = \gamma_{\text{SEP}}(\rho) \\
&= \gamma_{\text{SEP}}^*(\rho). \quad (44)
\end{aligned}$$

Recall that in this context, states should be understood as quantum channels with trivial input space, i.e., $\text{D}(AB) \equiv \text{CPTP}(\mathbb{C} \rightarrow AB)$. Note that the separable quantum channels with trivial input space are precisely equivalent to the set of separable states, i.e., $\text{SEP}(\mathbb{C} \otimes \mathbb{C} \rightarrow AB) \equiv \text{SEP}(A, B)$.

Proof. Considering the chain of inequalities in Fig. 3, it suffices to show that $\gamma_{\text{LO}}(\rho) \leq \gamma_{\text{SEP}}^*(\rho)$. For this purpose, it suffices to show that any QPD of ρ with respect to SEP^* induces a QPD of ρ with respect to LO with the identical ℓ_1 -norm of its coefficients.

For this purpose, note that $\text{LO}(\mathbb{C} \otimes \mathbb{C} \rightarrow AB)$ precisely corresponds to the set of states of the form $\tau \otimes \kappa$ where $\tau \in \text{D}(A)$, $\kappa \in \text{D}(B)$. Similarly, for γ_{SEP}^* , we have to consider the set $\text{SEP}^*(\mathbb{C} \otimes \mathbb{C} \rightarrow AB)$. By Lemma 3, it is given by

$$\begin{aligned}
\text{SEP}^*(\mathbb{C} \otimes \mathbb{C} \rightarrow AB) & \\
&\equiv \{(1-p)\sigma^+ - p\sigma^- : p \in [0, 1], \sigma^\pm \in \text{SEP}(A, B)\}. \quad (45)
\end{aligned}$$

Let $\rho = \sum_i a_i \sigma_i$ be a QPD where $\sigma_i \in \text{SEP}^*$, i.e., we can write $\sigma_i = (1-p_i)\sigma_i^+ - p_i\sigma_i^-$ for $p_i \in [0, 1]$, $\sigma_i^\pm \in \text{SEP}(A, B)$. Thanks to separability, we can write $\sigma_i^\pm = \sum_j q_j^{i,\pm} \tau_j^{i,\pm} \otimes \kappa_j^{i,\pm}$ where $(q_j^{i,\pm})_j$ is a probability distribution and $\tau_j^{i,\pm} \in \text{D}(A)$, $\kappa_j^{i,\pm} \in \text{D}(B)$. We can then rewrite

$$\rho = \sum_{i,j} a_i (1-p_i) q_j^{i,+} \tau_j^{i,+} \otimes \kappa_j^{i,+} - \sum_{i,j} a_i p_i q_j^{i,-} \tau_j^{i,-} \otimes \kappa_j^{i,-}, \quad (46)$$

which is a QPD with respect to $\text{LO}(\mathbb{C} \otimes \mathbb{C} \rightarrow AB)$ with ℓ_1 -norm of its coefficients given by

$$\sum_{i,j} |a_i (1-p_i) q_j^{i,+}| + |a_i p_i q_j^{i,-}| \leq \sum_i |a_i|. \quad (47)$$

The next result states that in the LO setting, entangling maps cannot be simulated at all without classical side information. Indeed, the only simulable maps in this setting are nonsignalling maps, i.e., maps that do not allow for the

⁸For LOCC one can use one round of classical communication to randomly choose between one of two protocols to perform.

transmission of any kind of information. In contrast, by using classical side information all Hermitian preserving superoperators can be achieved.⁹

Definition 4. A quantum operation $\mathcal{E} \in \text{HP}(AB \rightarrow A'B')$ is said to be nonsignalling if

$$\begin{aligned} \forall \rho, \sigma \in \text{D}(AB) \quad & \text{with} \\ \text{tr}_B[\rho] = \text{tr}_B[\sigma] \quad & \text{we have} \quad \text{tr}_B[\mathcal{E}(\rho)] = \text{tr}_B[\mathcal{E}(\sigma)]. \end{aligned} \quad (48)$$

$$\text{NSIG}(AB \rightarrow A'B') = \left\{ \sum_{i,j} \alpha_{ij} \mathcal{F}_i \otimes \mathcal{G}_j : \alpha_{ij} \in \mathbb{R}, \mathcal{F}_i \in \text{HPTP}(A \rightarrow A'), \mathcal{G}_j \in \text{HPTP}(B \rightarrow B') \right\}. \quad (49)$$

For a proof, we refer to Ref. [24, Theorem 13] or Ref. [25, Theorem 5.1]. Clearly, $\text{span}_{\mathbb{R}}\{\text{LO}(AB \rightarrow A'B')\}$ is contained in this set. To show the converse, it suffices to argue that any $\mathcal{F} \otimes \mathcal{G}$ for $\mathcal{F} \in \text{HPTP}(A \rightarrow A')$, $\mathcal{G} \in \text{HPTP}(B \rightarrow B')$ lies in $\text{span}_{\mathbb{R}}\{\text{LO}(AB \rightarrow A'B')\}$. This is a direct consequence of a slightly simpler result: Any HPTP map $\mathcal{E} \in \text{HPTP}(I \rightarrow O)$ can be decomposed into a QPD of two CPTP maps. Applying this to $\mathcal{F}$ and $\mathcal{G}$ provides the desired statement. One such valid QPD of $\mathcal{E}$ is given by

$$\mathcal{E} = d_{IO}\lambda\mathcal{D} - (d_{IO}\lambda - 1)\left(\frac{d_{IO}\lambda}{d_{IO}\lambda - 1}\mathcal{D} - \frac{1}{d_{IO}\lambda - 1}\mathcal{E}\right), \quad (50)$$

where $\mathcal{D}(\rho)$ is the fully depolarizing channel from I to O and $d_{IO} := \dim(I)\dim(O)$ and $\lambda \in \mathbb{R}$ is at least the maximal eigenvalue of $\mathcal{J}(\mathcal{E})$. Indeed, $(\frac{d_{IO}\lambda}{d_{IO}\lambda - 1}\mathcal{D} - \frac{1}{d_{IO}\lambda - 1}\mathcal{E})$ can easily be checked to be CPTP as

$$\begin{aligned} \text{tr}_O\left[\frac{d_{IO}\lambda}{d_{IO}\lambda - 1}\mathcal{J}(\mathcal{D}) - \frac{1}{d_{IO}\lambda - 1}\mathcal{J}(\mathcal{E})\right] \\ = \left(\frac{d_{IO}\lambda}{d_{IO}\lambda - 1} - \frac{1}{d_{IO}\lambda - 1}\right)\frac{1}{d_I}\mathbb{1}_I = \frac{1}{d_I}\mathbb{1}_I \end{aligned} \quad (51)$$

and

$$\left(\frac{d_{IO}\lambda}{d_{IO}\lambda - 1}\mathcal{J}(\mathcal{D}) - \frac{1}{d_{IO}\lambda - 1}\mathcal{J}(\mathcal{E})\right) \geq 0, \quad (52)$$

where the final step is equivalent to $d_{IO}\lambda \frac{1}{d_{IO}}\mathbb{1}_{IO} \geq \mathcal{J}(\mathcal{E})$.

Next, we consider the span of $\text{LO}^*(AB)$. Clearly, by definition we have

$$\text{span}_{\mathbb{R}}(\text{LO}^*(AB)) \subset \text{HP}(AB \rightarrow A'B'), \quad (53)$$

as any real linear combination of Hermitian preserving maps is itself Hermitian preserving, so it suffices to show the converse. For this purpose, notice that

$$\begin{aligned} \text{HP}(AB \rightarrow A'B') \\ = \left\{ \sum_{i,j} \mathcal{F}_i \otimes \mathcal{G}_j : \mathcal{F}_i \in \text{HP}(A \rightarrow A'), \mathcal{G}_j \in \text{HP}(B \rightarrow B') \right\}. \end{aligned} \quad (54)$$

We denote the set of such nonsignalling operation by $\text{NSIG}(AB \rightarrow A'B')$.

Proposition 2. The set of superoperators that exhibit a QPD under LO and LO^* are given by $\text{span}_{\mathbb{R}}(\text{LO}(AB \rightarrow A'B')) = \text{NSIG}(AB \rightarrow A'B')$ and $\text{span}_{\mathbb{R}}(\text{LO}^*(AB \rightarrow A'B')) = \text{HP}(AB \rightarrow A'B')$, respectively.

Proof. We start by characterizing the span of $\text{LO}(AB)$. It can be shown that

In the following, we will argue that any $\mathcal{E} \in \text{HP}(I \rightarrow O)$ can be written as a QPD of two elements in $\mathcal{L}[\text{QI}(I \rightarrow O)]$. Applying that to each $\mathcal{F}_i$ and $\mathcal{G}_j$ implies the desired statement. We can split the Choi representation of $\mathcal{E}$ into a difference of two positive operators $\Lambda^{\pm}$

$$\mathcal{J}(\mathcal{E}) = \Lambda^+ - \Lambda^- = \lambda^+ \frac{\Lambda^+}{\lambda^+} - \lambda^- \frac{\Lambda^-}{\lambda^-}, \quad (55)$$

where we denote by $\lambda^{\pm}$ the maximal eigenvalues of $\dim(I)\text{tr}_O[\Lambda^{\pm}]$. We can observe that $\Lambda^{\pm}/\lambda^{\pm}$ are Choi operators representing completely positive trace-nonincreasing maps, as they are positive and $\text{tr}_O[\Lambda^{\pm}/\lambda^{\pm}] \leq \dim(I)^{-1}\mathbb{1}_I$. The desired statement follows immediately from $\text{CPTN}(I \rightarrow O) \subset \mathcal{L}[\text{QI}(I \rightarrow O)]$. Note that we implicitly assumed that $\lambda^{\pm} > 0$, but the proof can be straightforwardly adapted for the case where $\mathcal{J}(\mathcal{E})$ is positive or negative. ■

Interestingly, the situation is different for the LOCC, SEP, and PPT settings. Indeed, here it is possible to simulate any trace-preserving map without using classical side information.

Proposition 3. The set of superoperators that exhibit a QPD under LOCC, LOCC*, SEP, SEP*, PPT, and PPT* are given by¹⁰

$$\text{span}_{\mathbb{R}}(\text{LOCC}) = \text{span}_{\mathbb{R}}(\text{SEP}) = \text{span}_{\mathbb{R}}(\text{PPT}) = \mathbb{R}\text{HPTP} \quad (56)$$

and

$$\text{span}_{\mathbb{R}}(\text{LOCC}^*) = \text{span}_{\mathbb{R}}(\text{SEP}^*) = \text{span}_{\mathbb{R}}(\text{PPT}^*) = \text{HP}, \quad (57)$$

where $\mathbb{R}\text{HPTP} := \{r \cdot \mathcal{E} : r \in \mathbb{R}, \mathcal{E} \in \text{HPTP}\}$.

Proof. We start by proving Eq. (57). Clearly, one has

$$\text{span}_{\mathbb{R}}(\text{LOCC}^*) \subset \text{span}_{\mathbb{R}}(\text{SEP}^*) \subset \text{span}_{\mathbb{R}}(\text{PPT}^*) \subset \text{HP}, \quad (58)$$

as any (real) linear combination of Hermitian preserving maps is itself Hermitian preserving. The converse is a direct consequence from Proposition 2, i.e.,

$$\text{HP} = \text{span}_{\mathbb{R}}(\text{LO}^*) \subset \text{span}_{\mathbb{R}}(\text{LOCC}^*). \quad (59)$$

⁹Note that $\text{CPTP}(AB \rightarrow A'B') \subset \text{HP}(AB \rightarrow A'B')$.

¹⁰Note that here we omit the input systems A, B and output systems A', B' to improve readability. For instance, LOCC stands for $\text{LOCC}(AB \rightarrow A'B')$.

Next, we prove Eq. (56). Again, one has

$$\text{span}_{\mathbb{R}}(\text{LOCC}) \subset \text{span}_{\mathbb{R}}(\text{SEP}) \subset \text{span}_{\mathbb{R}}(\text{PPT}) \subset \mathbb{RHPTP}, \quad (60)$$

as any (real) linear combination of trace-preserving maps is itself proportional to a trace-preserving map. It thus suffices to show $\text{HPTP} \subset \text{span}_{\mathbb{R}}(\text{LOCC})$, which then implies $\mathbb{RHPTP} \subset \text{span}_{\mathbb{R}}(\text{LOCC})$. We also know from the proof of Proposition 2 that any HPTP map can be decomposed as a QPD of CPTP maps. Hence, it suffices to show that any $\mathcal{E} \in \text{CPTP}(AB \rightarrow A'B')$ lies in $\text{span}_{\mathbb{R}}(\text{LOCC})$. We assume without loss of generality that $A = A'$ and $B = B'$ and that both are qubit systems (otherwise, we can always embed these systems into a large enough qubit system). Consider the Stinespring dilation of $\mathcal{E}$,

$$\mathcal{E}(\rho) = \text{tr}_E[U_{AEB}(\rho \otimes |0\rangle\langle 0|_E)U_{AEB}^\dagger]. \quad (61)$$

It suffices to show that the unitary channel $\mathcal{U} \in \text{CPTP}(AEB)$ induced by U_{AEB} can be decomposed into channels in $\text{LOCC}(AEB)$ (here the bipartition is chosen to be across AE and B). The unitary U_{AEB} can be decomposed into single-qubit gates and CNOT gates [26, Section 4.5.3]. All CNOT gates that act nonlocally between, e.g., AE and B , can be realized using CNOT gate teleportation [27], i.e., using LOCC and consuming a preshared entangled Bell pair,

$$\mathcal{U}(\rho_{AEB}) = \mathcal{F}(\rho_{AEB} \otimes |\Psi\rangle\langle\Psi|_{C_AC_B}), \quad (62)$$

where $\mathcal{F} \in \text{LOCC}(AEC_AC_BB)$ (here the bipartition is across AEC_A and C_BB) and $|\Psi\rangle\langle\Psi|_{C_AC_B}$ is a collection of Bell pairs between the ancilla systems C_A and C_B . Since there exists a QPD of the Bell pairs

$$|\Psi\rangle\langle\Psi|_{C_AC_B} = a^+\sigma^+ - a^-\sigma^-, \quad (63)$$

into separable states $\sigma^\pm \in \text{SEP}(C_AC_B)$ (see, e.g., Ref. [5,14]), we can thus decompose $\mathcal{U}$ into two $\text{LOCC}(AEB)$ maps

$$\mathcal{U}(\rho_{AEB}) = a^+\mathcal{F}(\rho_{AEB} \otimes \sigma_{C_AC_B}^+) - a^-\mathcal{F}(\rho_{AEB} \otimes \sigma_{C_AC_B}^-). \quad (64)$$

So any map that is trace-preserving (or proportional to a trace-preserving map) can be simulated without classical side information. Note that this does not mean that the extent with and without side information are equal—a priori the quasiprobability simulation with classical side information could be strictly more efficient. We next show that this is not the case in the SEP and PPT settings.

Proposition 4. Let $\mathcal{E}$ be a HPTP map. Then, $\gamma_{\text{PPT}}(\mathcal{E}) = \gamma_{\text{PPT}^*}(\mathcal{E})$ and $\gamma_{\text{SEP}}(\mathcal{E}) = \gamma_{\text{SEP}^*}(\mathcal{E})$.

Proposition 4 is a consequence of the following lemma.

Lemma 7. Let $\mathcal{Q}$ be a set of quantum instruments that fulfills the three properties in Definition 3 and $\forall(\mathcal{E}_i)_i \in \mathcal{Q}$ and for each element $\mathcal{E}_i$ that is proportional to a trace-preserving map, the rescaled map $\frac{1}{\text{tr}[\mathcal{E}_i]}\mathcal{E}_i$ is itself a (single-element) quantum instrument in $\mathcal{Q}$. Let $S := \{\sum_i \mathcal{E}_i : (\mathcal{E}_i)_i \in \mathcal{Q}\}$ and $S^* := \mathcal{L}[\mathcal{Q}]$. Then, $\gamma_S(\mathcal{E}) = \gamma_{S^*}(\mathcal{E})$ for all HPTP maps $\mathcal{E}$.

As discussed in the previous section, the quantum instruments of the sets PPT and SEP fulfill the three properties in Definition 3. Furthermore, an element of any SEP/PPT instrument has a separable/PPT Choi state and can hence be rescaled to a SEP/PPT channel. Therefore, both PPT and SEP fulfill the requirements of Lemma 7.

Note that LOCC also satisfies the three properties in Definition 3 but does not fulfill the rescaling property. In fact, while $\text{LOCC} \subsetneq \text{SEP}$, it is known that any channel in SEP (or in fact any separable instrument) can be *stochastically* realized with LOCC. More precisely, for any $\mathcal{E} \in \text{SEP}$ there exists some probability $p > 0$ such that $(p\mathcal{E}, (1-p)\mathcal{D}) \in \text{QI}_{\text{LOCC}}$ where $\mathcal{D}$ is the fully depolarizing channel [23,28].

We conclude that the technique from Lemma 7 is not applicable to the LOCC setting. However, this does not rule out the possibility that other LOCC constructions could exist which achieve the same extent as the LOCC^* construction.

Proof. The relation $S \subseteq S^*$ implies $\gamma_S(\mathcal{E}) \geq \gamma_{S^*}(\mathcal{E})$. Hence it suffices to show that given any QPD of $\mathcal{E}$ with respect to S^* , we can construct a QPD with respect to S with at most the same ℓ_1 -norm of its coefficients. Since $\mathcal{Q}$ is closed under mixture, we know by Lemmas 4 and 5 that both S and S^* are convex. Therefore, by Lemma 1 we can assume without loss of generality, that any QPD with respect to S^* has two elements.

Let

$$\mathcal{E} = a_+\mathcal{F}_+ - a_-\mathcal{F}_- \quad (65)$$

be a QPD with respect to S^* . By Lemma 3, $\mathcal{F}_\pm \in S^*$ can be written as

$$\mathcal{F}_\pm = \mathcal{G}_\pm^+ - \mathcal{G}_\pm^- \quad (66)$$

for some instruments $(\mathcal{G}_+^+, \mathcal{G}_+^-)$ and $(\mathcal{G}_-^+, \mathcal{G}_-^-)$ that lie in $\mathcal{Q}$. We can regroup this expression as

$$\begin{aligned} \mathcal{E} &= a_+\mathcal{F}_+ - a_-\mathcal{F}_- = a_+(\mathcal{G}_+^+ - \mathcal{G}_+^-) - a_-(\mathcal{G}_-^+ - \mathcal{G}_-^-) \\ &= (a_+\mathcal{G}_+^+ + a_-\mathcal{G}_-^-) - (a_+\mathcal{G}_+^- + a_-\mathcal{G}_-^+). \end{aligned} \quad (67)$$

Now we show that these two grouped terms are proportional to a valid map in S .

First, we show that they are proportional to a CPTP map. Being the sum of quantum instruments, complete positivity is clear, so it remains to show that the map is trace preserving. For this, we consider the reduced Choi state. By naming the target system of the channels $\mathcal{C}$, let us introduce

$$\mathcal{M} = \text{tr}_{\mathcal{C}}[\mathcal{J}(\mathcal{G}_+^+)] \quad \text{and} \quad \mathcal{N} = \text{tr}_{\mathcal{C}}[\mathcal{J}(\mathcal{G}_-^-)]. \quad (68)$$

Since the instruments fulfill $\mathcal{G}_\pm^+ + \mathcal{G}_\pm^- \in \text{CPTP}$, we find for the other two instruments

$$\text{tr}_{\mathcal{C}}[\mathcal{J}(\mathcal{G}_+^-)] = \frac{1}{d} - \mathcal{M} \quad \text{and} \quad \text{tr}_{\mathcal{C}}[\mathcal{J}(\mathcal{G}_-^+)] = \frac{1}{d} - \mathcal{N}, \quad (69)$$

where d is the dimension of the input system and $\frac{1}{d}$ describes the maximally mixed state. Applying the partial trace to Eq. (67), we get

$$\frac{1}{d} = a_+\mathcal{M} + a_-\mathcal{N} - \left(a_+\left(\frac{1}{d} - \mathcal{M}\right) + a_-\left(\frac{1}{d} - \mathcal{N}\right) \right), \quad (70)$$

which can be rewritten as

$$\frac{1 + a_+ + a_-}{2} \frac{1}{d} = a_+\mathcal{M} + a_-\mathcal{N}. \quad (71)$$

Hence, $\mathcal{P}_+ := \frac{2}{1+a_++a_-}(a_+\mathcal{G}_+^+ + a_-\mathcal{G}_-^-)$ is a CPTP map. Since $\mathcal{Q}$ is trivially fine-grainable, the following instrument is an

element of $\mathcal{Q}$:

$$\left(\frac{a_+}{a_+ + a_-} \mathcal{G}_+^+, \frac{a_+}{a_+ + a_-} \mathcal{G}_+^-, \frac{a_-}{a_+ + a_-} \mathcal{G}_-^+, \frac{a_-}{a_+ + a_-} \mathcal{G}_-^- \right). \quad (72)$$

Since $\mathcal{Q}$ is coarse-grainable, and $\mathcal{P}_+$ is proportional to $\frac{a_+}{a_+ + a_-} \mathcal{G}_+^+ + \frac{a_-}{a_+ + a_-} \mathcal{G}_-^+$, we can use the rescaling assumption to conclude that $\mathcal{P}_+$ is in S . Analogously, we find

$$\begin{aligned} \text{tr}_C[a_+ \mathcal{J}(\mathcal{G}_+^-) + a_- \mathcal{J}(\mathcal{G}_-^+)] \\ = a_+ \left(\frac{\mathbb{1}}{d} - \mathcal{M} \right) + a_- \left(\frac{\mathbb{1}}{d} - \mathcal{N} \right) \end{aligned} \quad (73)$$

$$= (a_+ + a_-) \frac{\mathbb{1}}{d} - (a_+ \mathcal{M} + a_- \mathcal{N}) \quad (74)$$

$$= (a_+ + a_-) \frac{\mathbb{1}}{d} - \frac{1 + a_+ + a_-}{2} \frac{\mathbb{1}}{d} \quad (75)$$

$$= \frac{a_+ + a_- - 1}{2} \frac{\mathbb{1}}{d}, \quad (76)$$

where the third equation uses Eq. (71). Thus, $\mathcal{P}_- := \frac{2}{a_+ + a_- - 1} (a_+ \mathcal{G}_+^- + a_- \mathcal{G}_-^+)$ is in S as well. With this, we found a

decomposition of $\mathcal{E}$ into maps in S of the form

$$\mathcal{E} = \frac{1 + a_+ + a_-}{2} \mathcal{P}_+ - \frac{a_+ + a_- - 1}{2} \mathcal{P}_- \quad (77)$$

and therefore

$$\gamma_S(\mathcal{E}) \leq \frac{1 + a_+ + a_-}{2} + \frac{a_+ + a_- - 1}{2} = a_+ + a_-. \quad (78)$$

■

VI. PPT LOWER BOUND

Proposition 4 implies that the quantity γ_{PPT} , which was introduced in Ref. [15], bounds not only γ_{LO} and γ_{LOCC} from below but also the physically more relevant quantities γ_{LO^*} and γ_{LOCC^*} , respectively. For any $\mathcal{E} \in \text{TP}(AB)$ we have

$$\gamma_{\text{LO}^*}(\mathcal{E}) \geq \gamma_{\text{LOCC}^*}(\mathcal{E}) \geq \gamma_{\text{PPT}^*}(\mathcal{E}) \stackrel{\text{Proposition 4}}{=} \gamma_{\text{PPT}}(\mathcal{E}). \quad (79)$$

This bound is interesting for two reasons:

(1) It can be evaluated efficiently via semidefinite programming [15, Eq. B2]. Let $\mathcal{J}(\mathcal{E})_{AA'BB'}$ denote the Choi state of $\mathcal{E} \in \text{HPTP}(AB \rightarrow A'B')$, then

$$\gamma_{\text{PPT}}(\mathcal{E}) = \min_{c \geq 0, X_{AA'BB'}} \{2c - 1 : X^{\text{T}_{BB'}} \geq 0, X^{\text{T}_{BB'}} \geq \mathcal{J}(\mathcal{E})^{\text{T}_{BB'}}, X_{AB} = c \mathbb{1}_{AB}, X \geq \mathcal{J}(\mathcal{E})\}, \quad (80)$$

which is a SDP.

(2) As we will see in the below, the lower bound from Eq. (79) is tight for many instances of $\mathcal{E}$.

The remainder of this section discusses the PPT lower bound from Eq. (79) and its consequences in more detail. First, we show that for the special case of pure states, the PPT lower bound is tight (recall that for states, $\gamma_{\text{LOCC}} = \gamma_{\text{LO}} = \gamma_{\text{SEP}}$).

Proposition 5. For any pure bipartite state $|\psi\rangle_{AB}$ one has $\gamma_{\text{SEP}}(|\psi\rangle\langle\psi|) = \gamma_{\text{PPT}}(|\psi\rangle\langle\psi|)$.

Proof. Since $\text{SEP} \subseteq \text{PPT}$ we have $\gamma_{\text{SEP}}(|\psi\rangle\langle\psi|) \geq \gamma_{\text{PPT}}(|\psi\rangle\langle\psi|)$. Hence, it remains to show the converse. Let us introduce a quantity analogous to the robustness of entanglement but relative to PPT instead of SEP, i.e.,

$$R_{\text{PPT}}(\rho_{AB}) := \min_{\sigma_{AB} \in \text{PPT}(AB), t \geq 0} \left\{ t : \frac{\rho_{AB} + t\sigma_{AB}}{1+t} \in \text{PPT} \right\}. \quad (81)$$

This quantity trivially fulfills $R_{\text{PPT}}(|\psi\rangle\langle\psi|) \leq R_{\text{SEP}}(|\psi\rangle\langle\psi|)$. In the following, we show the opposite direction, which implies the proposition by Lemma 2. To prove the desired statement, we must show that there exists no $\sigma \in S$, such that $t < R_{\text{SEP}}(|\psi\rangle\langle\psi|)$ and

$$\frac{1}{1+t} (|\psi\rangle\langle\psi|_{AB} + t\sigma_{AB}) \in \text{PPT}. \quad (82)$$

We follow the proof in Ref. [14] and extend it from separable to PPT states. Given a σ that fulfills Eq. (82), we show that t has to be larger or equal to $R_{\text{SEP}}(|\psi\rangle\langle\psi|)$. Using the property of PPT states, one obtains for any $|\phi\rangle$

$$\begin{aligned} \langle\phi| \frac{1}{1+t} (|\psi\rangle\langle\psi|_{AB} + t\sigma_{AB})^{\text{T}_B} |\phi\rangle &\geq 0 \quad \text{and} \\ \langle\phi| (|\psi\rangle\langle\psi|_{AB}^{\text{T}_B} + t\sigma_{AB}^{\text{T}_B}) |\phi\rangle &\geq 0. \end{aligned} \quad (83)$$

In the following, we want to work in the Schmidt basis $|\phi\rangle = \sum_i a_i |i\rangle_A |i\rangle_B$. Using this basis, we now define Bell states

$$|\phi_{ij}^+\rangle = \frac{1}{\sqrt{2}}(|ij\rangle + |ji\rangle) \quad \text{and} \quad |\phi_{ij}^-\rangle = \frac{1}{\sqrt{2}}(|ij\rangle - |ji\rangle) \quad (84)$$

and corresponding projectors

$$P_{ij}^+ = |\phi_{ij}^+\rangle\langle\phi_{ij}^+| \quad \text{and} \quad P_{ij} = |\phi_{ij}^-\rangle\langle\phi_{ij}^-|. \quad (85)$$

This allows us to write

$$|\psi\rangle\langle\psi|^{\text{T}_B} = \left(\sum_i a_i |ii\rangle \sum_j a_j \langle jj| \right)^{\text{T}_B} \quad (86)$$

$$= \left(\sum_i a_i^2 |ii\rangle\langle ii| + \sum_{i \neq j} a_i a_j |ii\rangle\langle jj| \right)^{\text{T}_B} \quad (87)$$

$$= \sum_i a_i^2 |ii\rangle\langle ii| + \sum_{i \neq j} a_i a_j |ij\rangle\langle ji| \quad (88)$$

$$= \sum_i a_i^2 |ii\rangle\langle ii| + \sum_{i > j} a_i a_j (|ij\rangle\langle ji| + |ji\rangle\langle ij|) \quad (89)$$

$$= \sum_i a_i^2 |ii\rangle\langle ii| + \sum_{i > j} a_i a_j (P_{ij}^+ - P_{ij}). \quad (90)$$

Since Eq. (83) holds for any $|\phi\rangle$, we can pick $|\phi\rangle = |\phi_{ij}^-\rangle$. This leads to

$$\langle\phi_{ij}^-| (-a_i a_j P_{ij} + t\sigma^{\text{T}_B}) |\phi_{ij}^- \rangle \geq 0 \quad (91)$$

or equivalently

$$\text{tr}[P_{ij} \sigma^{\text{T}_B}] \geq \frac{a_i a_j}{t}. \quad (92)$$

From Ref. [14], we know that the robustness of entanglement can be written as

$$\begin{aligned} R_{\text{SEP}}(|\psi\rangle) &= \left(\sum_i a_i \right)^2 - 1 = \sum_{ij} a_i a_j - 1 \\ &= \sum_{ij} a_i a_j - \sum_i a_i^2 = 2 \sum_{i>j} a_i a_j. \end{aligned} \quad (93)$$

By summing over Eq. (92), we then find

$$\text{tr} \left[\sum_{i>j} P_{ij} \sigma^{T_B} \right] \geq \frac{R_{\text{SEP}}(|\psi\rangle)}{2t}. \quad (94)$$

Thus, if we can show that $\text{tr}[\sum_{i>j} P_{ij} \sigma^{T_B}] \leq 1/2$, then the statement is shown. Let us define $M := \sum_{i>j} P_{ij}$, then using cyclicity of the trace, we have

$$\text{tr}[M \sigma^{T_B}] = \text{tr}[M^{T_B} \sigma] \leq \lambda_{\max}(M^{T_B}), \quad (95)$$

where $\lambda_{\max}(M^{T_B})$ denotes the largest eigenvalue of M^{T_B} . To bound this quantity we write

$$\begin{aligned} M^{T_B} &= \left(\sum_{i>j} P_{ij} = \frac{1}{2} \sum_{i>j} |ij\rangle\langle ij| \right. \\ &\quad \left. + |ji\rangle\langle ji| - |ij\rangle\langle ji| - |ji\rangle\langle ij| \right)^{T_B} \end{aligned} \quad (96)$$

$$= \frac{1}{2} \sum_{i>j} |ij\rangle\langle ij| + |ji\rangle\langle ji| - |ii\rangle\langle jj| - |jj\rangle\langle ii| \quad (97)$$

$$= \frac{1}{2} \sum_{i \neq j} |ij\rangle\langle ij| - |ii\rangle\langle jj|. \quad (98)$$

In this form, it is easy to see that $M^{T_B} \leq \frac{1}{2} \mathbb{1}$, since

$$\frac{1}{2} \mathbb{1} - M^{T_B} = \frac{1}{2} \sum_{ij} |ij\rangle\langle ij| - \frac{1}{2} \sum_{i \neq j} |ij\rangle\langle ij| + \frac{1}{2} \sum_{i \neq j} |ii\rangle\langle jj| \quad (99)$$

$$= \frac{1}{2} \left(\sum_i |ii\rangle\langle ii| + \sum_{i \neq j} |ii\rangle\langle jj| \right) \quad (100)$$

$$= \frac{1}{2} \left(\sum_i |ii\rangle \right) \left(\sum_j \langle jj| \right), \quad (101)$$

which is proportional to a rank 1 projector and therefore greater equal 0. Therefore the largest eigenvalue can be bounded by 1/2 and we get

$$t \geq R_{\text{SEP}}(|\psi\rangle), \quad (102)$$

which concludes the proof. ■

In previous work (see Refs. [5,6]), the main technical tool to find lower bounds for γ_{LO^*} and γ_{LOCC^*} was not based on γ_{PPT} but on a different approach. Consider any bipartite channel $\mathcal{E} \in \text{CPTP}(AB \rightarrow A'B')$ and any separable state $\sigma \in \text{SEP}(A\bar{A}, B\bar{B})$ for some ancillary systems $\bar{A}, \bar{B}$. Then one has

$$\gamma_{\text{LOCC}^*}(\mathcal{E}) \geq \gamma_{\text{SEP}}(\mathcal{E}(\sigma)), \quad (103)$$

since any QPD of $\mathcal{E}$ directly induces a QPD of $\mathcal{E}(\sigma)$ with same ℓ_1 -norm of its coefficients. Typically, σ is chosen to be the product of maximally entangled states across $A, \bar{A}$ and $B, \bar{B}$, respectively. The resulting state $\mathcal{E}(\sigma)$ is then the Choi state of $\mathcal{E}$.

The following example illustrates that $\gamma_{\text{PPT}}(\mathcal{U})$ of a unitary channel $\mathcal{U}$ can be a strictly better lower bound than using a lower bound based on the Choi state.

Example 2. Consider the Toffoli gate. There are two ways (up to permutation of control qubits) how to group the three qubits into two bipartitions. By solving the SDP given in Eq. (80) we find for any bipartition that

$$\begin{aligned} \gamma_{\text{PPT}}(\text{Toffoli}) &= 3 > \gamma_{\text{SEP}}(\mathcal{J}(\text{Toffoli})) \\ &= \frac{1}{4}(\sqrt{6} + \sqrt{2})^2 - 1 \approx 2.73, \end{aligned} \quad (104)$$

where the penultimate step follows from Lemma 2 and Eq. (93). It has been shown in Ref. [6] that $\gamma_{\text{LO}^*}(\text{Toffoli}) = 3$, and therefore the PPT lower bound turns out to be tight in this case.

As another consequence of Proposition 5, one can also show that the PPT lower bound is tight for a large class of bipartite unitary channels.

Corollary 1. Consider a bipartite unitary of the form

$$U_{AB} = (V_A^{(1)} \otimes V_B^{(2)}) \sum_k u_k (L_k)_A \otimes (R_k)_B (V_A^{(3)} \otimes V_B^{(4)}), \quad (105)$$

where $V_A^{(1)}, V_B^{(2)}, V_A^{(3)}, V_B^{(4)}, (L_k)_A$, and $(R_k)_B$ are unitaries on A and B such that the $(L_k)_k$ and $(R_k)_k$ form orthogonal sets under the Hilbert-Schmidt inner product. Denote the channel induced by U_{AB} by $\mathcal{U}(\rho) := U \rho U^\dagger$. Then

$$\gamma_{\text{PPT}}(\mathcal{U}) = \gamma_{\text{LOCC}^*}(\mathcal{U}) = \gamma_{\text{LO}^*}(\mathcal{U}) = 1 + 2 \sum_{k \neq k'} |u_k| |u_{k'}|. \quad (106)$$

Proof. This is a consequence of Ref. [6, Theorem 5.1]. By simple considerations of hierarchies in the decomposition sets as well as application of the Choi bound and Proposition 5, one directly obtains

$$\begin{aligned} \gamma_{\text{LO}^*}(\mathcal{U}) &\geq \gamma_{\text{LOCC}^*}(\mathcal{U}) \geq \gamma_{\text{PPT}^*}(\mathcal{U}) \geq \gamma_{\text{PPT}}(\mathcal{J}(\mathcal{U})) \\ &= \gamma_{\text{SEP}}(\mathcal{J}(\mathcal{U})). \end{aligned} \quad (107)$$

The result [6, Theorem 5.1] then shows that $\gamma_{\text{LO}^*}(\mathcal{U}) = 1 + 2 \sum_{k \neq k'} |u_k| |u_{k'}| = \gamma_{\text{SEP}}(\mathcal{J}(\mathcal{U}))$, which concludes the proof.

Notice that the unitaries described in Corollary 1 contain a large class of unitaries that are used in practice. Any two-qubit unitary (and tensor products thereof) has the above form due to Cartan's KAK decomposition [29].

An interesting consequence of Proposition 5 is that for unitary channels the lower bound $\gamma_{\text{LOCC}^*} \geq \gamma_{\text{PPT}}$ is at least as good as state-based lower bounds, even if one optimizes over the state.

Corollary 2. For any bipartite unitary channel $\mathcal{U} \in \text{CPTP}(AB)$ one has

$$\gamma_{\text{LOCC}^*}(\mathcal{U}) \geq \gamma_{\text{PPT}}(\mathcal{U}) \geq \sup_{\sigma \in \mathcal{D}(A\bar{A}B\bar{B}')} \frac{\gamma_{\text{SEP}}(\mathcal{U}(\sigma))}{\gamma_{\text{SEP}}(\sigma)}, \quad (108)$$

where A' and B' are arbitrarily large ancillary systems.

Proof. The first inequality in the assertion of Corollary 2 follows by noting that

$$\gamma_{\text{LOCC}^*}(\mathcal{U}) \geq \gamma_{\text{PPT}^*}(\mathcal{U}) \stackrel{\text{Proposition 4}}{=} \gamma_{\text{PPT}}(\mathcal{U}). \quad (109)$$

It thus remains to prove the second inequality of Corollary 2. For any choice of $\sigma \in D(AA'BB')$, we have

$$\gamma_{\text{LOCC}^*}(\mathcal{U})\gamma_{\text{LOCC}^*}(\sigma) \geq \gamma_{\text{LOCC}^*}(\mathcal{U}(\sigma)), \quad (110)$$

since two QPD for $\mathcal{U}$ and σ can be combined into a QPD for $\mathcal{U}(\sigma)$ with ℓ_1 -norm being the product of the ℓ_1 -norms of the original QPDs. In Ref. [30, Theorem 1], it was shown that any logarithmic decomposition-based entanglement monotone attains the worst-case on the set of free states. Concretely, this implies

$$\sup_{\sigma \in D(AA'BB')} \frac{\gamma_{\text{LOCC}^*}(\mathcal{U}(\sigma))}{\gamma_{\text{LOCC}^*}(\sigma)} = \sup_{\rho \in \text{SEP}(AA', BB')} \gamma_{\text{LOCC}^*}(\mathcal{U}(\rho)). \quad (111)$$

Additionally, we have $\gamma_{\text{LOCC}^*}(\rho) = \gamma_{\text{SEP}}(\rho)$ for states, as explained in Lemma 6. We then find

$$\begin{aligned} & \sup_{\sigma \in D(AA'BB')} \frac{\gamma_{\text{LOCC}^*}(\mathcal{U}(\sigma))}{\gamma_{\text{LOCC}^*}(\sigma)} \\ & \stackrel{\text{Eq. (111)}}{=} \sup_{\rho \in \text{SEP}(AA', BB')} \gamma_{\text{LOCC}^*}(\mathcal{U}(\rho)) \end{aligned} \quad (112)$$

$$\stackrel{\text{Lemma 6}}{=} \sup_{\rho \in \text{SEP}(AA', BB')} \gamma_{\text{SEP}}(\mathcal{U}(\rho)) \quad (113)$$

$$= \sup_{\psi \in \text{SEP}(AA', BB')} \gamma_{\text{SEP}}(\mathcal{U}(|\psi\rangle\langle\psi|)), \quad (114)$$

where the final step uses that γ_{SEP} is a convex quantity, and optimization of it over all states reduces to an optimization over all pure states. Using that $\gamma_{\text{SEP}}(|\psi\rangle\langle\psi|) = 1$ for any $\psi \in \text{SEP}$ yields

$$\begin{aligned} & \sup_{\psi \in \text{SEP}(AA', BB')} \gamma_{\text{SEP}}(\mathcal{U}(|\psi\rangle\langle\psi|)) \\ & = \sup_{\psi \in \text{SEP}(AA', BB')} \frac{\gamma_{\text{SEP}}(\mathcal{U}(|\psi\rangle\langle\psi|))}{\gamma_{\text{SEP}}(|\psi\rangle\langle\psi|)} \end{aligned} \quad (115)$$

$$\leq \sup_{\psi \in \overline{D}(AA'BB')} \frac{\gamma_{\text{SEP}}(\mathcal{U}(|\psi\rangle\langle\psi|))}{\gamma_{\text{SEP}}(|\psi\rangle\langle\psi|)} \quad (116)$$

$$\stackrel{\text{Proposition 5}}{=} \sup_{\psi \in \overline{D}(AA'BB')} \frac{\gamma_{\text{PPT}}(\mathcal{U}(|\psi\rangle\langle\psi|))}{\gamma_{\text{PPT}}(|\psi\rangle\langle\psi|)} \quad (117)$$

$$\leq \sup_{\sigma \in D(AA'BB')} \frac{\gamma_{\text{PPT}}(\mathcal{U}(\sigma))}{\gamma_{\text{PPT}}(\sigma)} \quad (118)$$

$$\stackrel{\text{Eq. (110)}}{\leq} \gamma_{\text{PPT}}(\mathcal{U}). \quad (119)$$

■

VII. DISCUSSION

In this work, we made the interesting observation that classical side information is absolutely crucial to realize circuit cutting without classical communication, whereas it provides no advantage in the SEP and PPT setting. This raises the following question: Does classical side information help to reduce the quasiprobability extent in the case of LOCC? In mathematical terms, do there exist channels $\mathcal{E}$

such that $\gamma_{\text{LOCC}^*}(\mathcal{E}) < \gamma_{\text{LOCC}}(\mathcal{E})$? As noted in the discussion of Lemma 7, our current techniques do not suffice to answer this question. Due to the notoriously complicated structure of LOCC, this problem could be challenging to resolve.

We do note that for some nontrivial families of channels, it is known that $\gamma_{\text{LOCC}^*} = \gamma_{\text{LOCC}}$. For instance, it has been shown in Ref. [5] that this is the case for Clifford unitaries. On the other hand, the optimal LOCC* QPDs for general two-qubit unitaries introduced in Ref. [6] heavily utilize the classical side information, and it seems unclear how the optimal overhead could also be achieved without it. This could be seen as evidence for a separation between γ_{LOCC^*} and γ_{LOCC} .

More broadly, it remains an important task for future work to better understand the quasiprobability extent of unitary channels. For unitaries, the quasiprobability extent $\gamma_{\text{LOCC}^*}(\mathcal{U})$ is currently only known in two cases: for Clifford gates [5] and for unitaries with Schmidt-like decompositions (including two-qubit unitaries) [6,7]. In both cases, our work shows that the PPT lower bound is tight, i.e., $\gamma_{\text{LOCC}^*}(\mathcal{U}) = \gamma_{\text{PPT}}(\mathcal{U})$.¹¹ It is natural to ask whether this equality also holds for general bipartite unitaries. Such a result would lift the statement of Proposition 5 from the setting of states to channels. From a more practical side, this would allow us to efficiently evaluate the quasiprobability extent for arbitrary unitary channels, which is relevant for circuit cutting applications.

ACKNOWLEDGMENTS

C.P. and L.S. acknowledge support from the ETH Quantum Center and the NCCR SwissMAP.

DATA AVAILABILITY

No data were created or analyzed in this study.

APPENDIX: EXTENT SEPARATIONS

In this section, we show that the quasiprobability extents of different decomposition sets differ as depicted by the horizontal inequalities in Fig. 1. In technical term we show that

$$\gamma_{\text{LO}} > \gamma_{\text{LOCC}^*} > \gamma_{\text{SEP}^*} > \gamma_{\text{PPT}^*} \quad (\text{A1})$$

and

$$\gamma_{\text{LO}} > \gamma_{\text{LOCC}} > \gamma_{\text{SEP}} > \gamma_{\text{PPT}}, \quad (\text{A2})$$

where $\gamma_{S_1} > \gamma_{S_2}$ means that there exists a channel $\mathcal{E}$ such that $\gamma_{S_1}(\mathcal{E}) > \gamma_{S_2}(\mathcal{E})$. This is a direct consequence of the following result.

Lemma A1. Consider two sets of quantum instruments $\mathcal{Q}_1 \subset \text{QI}(I \rightarrow O)$, $\mathcal{Q}_2 \subset \text{QI}(I \rightarrow O)$ and denote the induced sets of channels by $S_i := \{\sum_j \mathcal{E}_j : (\mathcal{E}_j)_j \in \mathcal{Q}_i\}$ and $S_i^* := \mathcal{L}[\mathcal{Q}_i]$ for $i = 1, 2$. Furthermore, assume that $\text{conv}(\overline{S_1}) \subsetneq S_2$ where $\text{conv}(\overline{S_1})$ denotes the closure of the convex hull of S_1 . Then there exists a channel $\mathcal{E} \in \text{CPTP}(I \rightarrow O)$ such that

$$\gamma_{S_2}(\mathcal{E}) = \gamma_{S_2^*}(\mathcal{E}) = 1 \quad \text{and} \quad \gamma_{S_1}(\mathcal{E}), \gamma_{S_1^*}(\mathcal{E}) > 1. \quad (\text{A3})$$

¹¹For Clifford gates, the statement follows from Proposition 5, Corollary 2, and $\gamma_{\text{LOCC}^*}(\mathcal{U}) = \gamma_{\text{SEP}}(\mathcal{J}(\mathcal{U}))$, which is proven in Ref. [5, Theorem 5.1].

Our desired statement in Eqs. (A1) and (A2) is a direct consequence of the above lemma by appropriately choosing Q_1, Q_2 to be $Q_{\text{LO}}, Q_{\text{LOCC}}, Q_{\text{LOCC}}, Q_{\text{SEP}},$ or $Q_{\text{SEP}}, Q_{\text{PPT}}$. Indeed, it is known that $\text{conv}(\text{LO})$ (which contains local operations with shared randomness) is a strict subset of LOCC , the closure LOCC is a strict subset of SEP (this has been demonstrated in the context of state discrimination [31–33] and random distillation [34,35]), and that SEP (which is convex and closed) is a strict subset of PPT (see, e.g., Ref. [36]).

Proof. Let us pick some $\mathcal{E}$ that lies in S_2 but not in $\text{conv}(S_1)$. Clearly, $\gamma_{S_2}(\mathcal{E}) = \gamma_{S_2^*}(\mathcal{E}) = 1$. We will now prove by contradiction that $\gamma_{S_1^*}(\mathcal{E})$ must be > 1 . This in turn will also imply that $\gamma_{S_1}(\mathcal{E}) > 1$, since $S_1 \subset S_1^*$.

Let us assume that $\gamma_{S_1^*}(\mathcal{E}) = 1$. Therefore, for any $\epsilon > 0$ there exists a QPD $\mathcal{E} = \sum_i a_i \mathcal{F}_i$ with $\mathcal{F}_i \in S_1^*$ and $\|a\|_1 \leq 1 + \epsilon$. We can write $\mathcal{F}_i = \sum_j b_j^{(i)} \mathcal{G}_j^{(i)}$ with $b_j^{(i)} \in [-1, 1]$ and $(\mathcal{G}_j^{(i)})_j \in Q_1$.

Now, define the superoperator $\tilde{\mathcal{E}} := \sum_{i,j} |a_i| \mathcal{G}_j^{(i)}$. For the Choi states of $\mathcal{E}$ and $\tilde{\mathcal{E}}$ it holds that

$$\mathcal{J}(\mathcal{E}) = \sum_{i,j} a_i b_j^{(i)} \mathcal{J}(\mathcal{G}_j^{(i)}) \leq \sum_{i,j} |a_i| \mathcal{J}(\mathcal{G}_j^{(i)}) = \mathcal{J}(\tilde{\mathcal{E}}). \quad (\text{A4})$$

Hence, the trace distance between these two Choi states is

$$\begin{aligned} \|\mathcal{J}(\mathcal{E}) - \mathcal{J}(\tilde{\mathcal{E}})\|_1 &= \text{tr}[\mathcal{J}(\tilde{\mathcal{E}}) - \mathcal{J}(\mathcal{E})] \\ &= \sum_{i,j} |a_i| \text{tr}[\mathcal{J}(\mathcal{G}_j^{(i)})] - 1 \\ &= \sum_i |a_i| - 1 \leq \epsilon. \end{aligned} \quad (\text{A5})$$

The channel $\frac{1}{\|a\|_1} \tilde{\mathcal{E}}$ lies in $\text{conv}(S_1)$ and approximates $\mathcal{E}$ up to error

$$\begin{aligned} \left\| \mathcal{J}(\mathcal{E}) - \frac{1}{\|a\|_1} \mathcal{J}(\tilde{\mathcal{E}}) \right\|_1 &\leq \|\mathcal{J}(\mathcal{E}) - \mathcal{J}(\tilde{\mathcal{E}})\|_1 \\ &\quad + \left\| \mathcal{J}(\tilde{\mathcal{E}}) - \frac{1}{\|a\|_1} \mathcal{J}(\tilde{\mathcal{E}}) \right\|_1 \\ &\leq \epsilon + \frac{\epsilon}{1 + \epsilon} \leq 2\epsilon. \end{aligned} \quad (\text{A6})$$

In summary, we have shown that our map $\mathcal{E}$ can be approximated arbitrarily well by elements in $\text{conv}(S_1)$. This implies that $\mathcal{E}$ lies in the closure of $\text{conv}(S_1)$, which is a contradiction. ■

-
- [1] S. Bravyi, G. Smith, and J. A. Smolin, Trading classical and quantum computational resources, *Phys. Rev. X* **6**, 021043 (2016).
 - [2] T. Peng, A. W. Harrow, M. Ozols, and X. Wu, Simulating large quantum circuits on a small quantum computer, *Phys. Rev. Lett.* **125**, 150504 (2020).
 - [3] K. Mitarai and K. Fujii, Constructing a virtual two-qubit gate by sampling single-qubit operations, *New J. Phys.* **23**, 023021 (2021).
 - [4] K. Mitarai and K. Fujii, Overhead for simulating a non-local channel with local channels by quasiprobability sampling, *Quantum* **5**, 388 (2021).
 - [5] C. Piveteau and D. Sutter, Circuit knitting with classical communication, *IEEE Trans. Inf. Theory* **70**, 2734 (2024).
 - [6] L. Schmitt, C. Piveteau, and D. Sutter, Cutting circuits with multiple two-qubit unitaries, *Quantum* **9**, 1634 (2025).
 - [7] A. W. Harrow and A. Lowe, Optimal quantum circuit cuts with application to clustered Hamiltonian simulation, *PRX Quantum* **6**, 010316 (2025).
 - [8] A. Carrera Vazquez, C. Tornow, D. Ristè, S. Woerner, M. Takita, and D. J. Egger, Combining quantum processors with real-time classical communication, *Nature (London)* **636**, 75 (2024).
 - [9] 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).
 - [10] L. Brenner, C. Piveteau, and D. Sutter, Optimal wire cutting with classical communication, *arXiv:2302.03366*.
 - [11] C. Piveteau, *Simulating Quantum Circuits with Restricted Quantum Computers*, Ph.D. thesis, ETH Zurich, 2025.
 - [12] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, *Phys. Rev. X* **8**, 031027 (2018).
 - [13] S. Gharibian, Strong NP-hardness of the quantum separability problem, *Quantum Info. Comput.* **10**, 343 (2010).
 - [14] G. Vidal and R. Tarrach, Robustness of entanglement, *Phys. Rev. A* **59**, 141 (1999).
 - [15] M. Jing, C. Zhu, and X. Wang, Circuit knitting facing exponential sampling-overhead scaling bounded by entanglement cost, *Phys. Rev. A* **111**, 012433 (2025).
 - [16] X. Zhao, L. Zhang, B. Zhao, and X. Wang, Power of quantum measurement in simulating unphysical operations, *Phys. Rev. Res.* **7**, 013334 (2025).
 - [17] A. Peres, Separability criterion for density matrices, *Phys. Rev. Lett.* **77**, 1413 (1996).
 - [18] M. Horodecki, P. Horodecki, and R. Horodecki, Separability of mixed states: Necessary and sufficient conditions, *Phys. Lett. A* **223**, 1 (1996).
 - [19] H. Pashayan, J. J. Wallman, and S. D. Bartlett, Estimating outcome probabilities of quantum circuits using quasiprobabilities, *Phys. Rev. Lett.* **115**, 070501 (2015).
 - [20] M. Howard and E. Campbell, Application of a resource theory for magic states to fault-tolerant quantum computing, *Phys. Rev. Lett.* **118**, 090501 (2017).
 - [21] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, *Phys. Rev. Lett.* **119**, 180509 (2017).
 - [22] M. Heinrich and D. Gross, Robustness of magic and symmetries of the stabiliser polytope, *Quantum* **3**, 132 (2019).
 - [23] E. Chitambar, D. Leung, L. Mancinska, M. Ozols, and A. Winter, Everything you always wanted to know about LOCC (but were afraid to ask), *Commun. Math. Phys.* **328**, 303 (2014).
 - [24] G. Gutoski, Properties of local quantum operations with shared entanglement, *Quantum Info. Comput.* **9**, 739 (2009).
 - [25] P. J. Cavalcanti, J. H. Selby, J. Sikora, and A. B. Sainz, Decomposing all multipartite nonsignalling channels via quasiprobabilistic mixtures of local channels in generalised

- probabilistic theories, *J. Phys. A: Math. Theor.* **55**, 404001 (2022).
- [26] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000).
- [27] D. Gottesman and I. L. Chuang, Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations, *Nature (London)* **402**, 390 (1999).
- [28] W. Dür, G. Vidal, and J. I. Cirac, Three qubits can be entangled in two inequivalent ways, *Phys. Rev. A* **62**, 062314 (2000).
- [29] N. Khaneja and S. J. Glaser, Cartan decomposition of $SU(2^n)$ and control of spin systems, *Chem. Phys.* **267**, 11 (2001).
- [30] E. T. Campbell, Optimal entangling capacity of dynamical processes, *Phys. Rev. A* **82**, 042314 (2010).
- [31] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters, Quantum nonlocality without entanglement, *Phys. Rev. A* **59**, 1070 (1999).
- [32] M. Koashi, F. Takenaga, T. Yamamoto, and N. Imoto, Quantum nonlocality without entanglement in a pair of qubits, [arXiv:0709.3196](https://arxiv.org/abs/0709.3196).
- [33] A. M. Childs, D. Leung, L. Mančinska, and M. Ozols, A framework for bounding nonlocality of state discrimination, *Commun. Math. Phys.* **323**, 1121 (2013).
- [34] W. Cui, E. Chitambar, and H.-K. Lo, Randomly distilling w -class states into general configurations of two-party entanglement, *Phys. Rev. A* **84**, 052301 (2011).
- [35] E. Chitambar, W. Cui, and H.-K. Lo, Increasing entanglement monotones by separable operations, *Phys. Rev. Lett.* **108**, 240504 (2012).
- [36] C. H. Bennett, D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin, and B. M. Terhal, Unextendible product bases and bound entanglement, *Phys. Rev. Lett.* **82**, 5385 (1999).