Reexamination of quantum state transformations with zero communication

Ian George and Eric Chitambar

Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

It is known that general convertibility of bipartite entangled states is not possible to arbitrary error without some classical communication. While some tradeoffs between communication cost and conversion error have been proven, these bounds can be very loose. In particular, there are many cases in which tolerable error might be achievable using zero-communication protocols. In this work we address these cases by deriving the optimal fidelity of pure quantum state conversions under local unitaries as well as local operations and shared randomness. We also use these results to explore catalytic conversions between pure states using zero communication.

I. INTRODUCTION

The theory of quantum mechanics through the lens of information and vice versa [1–3] has afforded the physicist and the information scientist alike with a new way to view the objects and long-term goals of their study. No better example of this can be found than quantum resource theories. Quantum resource theories specify the relevant physical property in such a manner as to better tease apart the complexities of quantum mechanics while also establishing what tasks may be achieved with said resource [4]. Perhaps the earliest example of such a resource theory is the resource theory of entanglement. Entanglement may be viewed as a form of correlation that does not exist in the classical world [5]. Roughly speaking, the resource theory of entanglement asks (1) what tasks may be performed better using entangled states and (2) how entangled states may be converted from one to another under some class of free operations.

The most standard view of the resource theory of entanglement considers the set of free operations to be local operations and classical communication (LOCC) which captures the "distant laboratory" paradigm where two (or more) parties share an entangled state in spatially separated labs and they can only perform operations on their respective portions and exchange classical information (see Fig. 1). Not only is this the most standard set of free operations, but in some respect it seems minimal. Indeed, Hayden and Winter showed that to convert one (pure) entangled state to another to sufficiently small precision requires a certain amount of communication between labs, regardless of how many auxiliary electron paramagnetic resonance (EPR) pairs they share [6] (see also [7]). This is distinct not only from the classical setting [8], but also from quantum states that are not entangled [9,10]. However, the results of Hayden and Winter, while fundamental, do not give us a complete picture of the tradeoff between communication and achievable tolerated error in quantum state conversions. Indeed, it is easy to find examples of state conversions which, according to the best known lower bounds, still may be possible to perform with a tolerated error of 1% using no communication (see Example 1 of Sec. III).

This shows that a relatively large gap in our understanding of zero-communication entanglement transformations still persists, and one we aim to address in this work.

Moreover, the tools we develop to address this problem will also allow us to study quantum state transformations using shared auxiliary entanglement. The operational paradigm in which parties are allowed to use arbitrary preshared entanglement but no communication is known as local operations and shared entanglement (LOSE) [11]. Without restriction on the entangled resource, quantum state convertibility $|\psi\rangle_{AB} \rightarrow |\phi\rangle_{AB}$ under LOSE is trivial since Alice and Bob could always just demand $|\phi\rangle$ as their preshared entanglement and then throw away $|\psi\rangle$ when it is given. However, if one tries to minimize the amount of preshared entanglement (under some measure), the problem becomes quite interesting. A more demanding setting is to minimize the amount of preshared entanglement $|\omega\rangle$ while requiring it outputs the target state $|\phi\rangle$, i.e., $|\psi\rangle_{AB} \otimes |\omega\rangle_{A'B'} \rightarrow |\phi\rangle_{AB} \otimes |\omega\rangle_{A'B'}$ for auxiliary pre-shared entanglement $|\omega\rangle_{A'B'}$. Transformations of this form are known as catalytic transformations with $|\omega\rangle_{A'B}$ being the catalyst (see [12,13] for reviews on catalytic transformations).

When instead the catalyst only needs to be returned approximately, we distinguish it from the zero-error case by calling it an "embezzler" or "embezzling state." Remarkably, van Dam and Hayden showed that there exists a family of "universal embezzling states" under local operations [14]. That is, they found a family of states $\{|\mu(n)\rangle_{A'B'}\}_{n \in \mathbb{N}}$ such that for any tolerated nonzero error ε, one can always prepare a pure quantum state $|\psi\rangle_{AB}$ using a member of this family and zero communication. Perhaps even more surprising, the authors showed that this family of embezzlers is "optimal" in a certain sense. In particular, they showed that if one strengthens the class of transformations to local operations and classical communication (LOCC) and replaces the appropriate universal embezzler state $|\mu(n_\varepsilon)\rangle$ for some $n_\varepsilon \in \mathbb{N}$ to a state-dependent embezzler state $|\zeta(\psi, \varepsilon)\rangle$ that also depends on the tolerated error ε, the dimension of the entanglement in $|\mu(n_\varepsilon)\rangle$ and $|\zeta(\psi, \varepsilon)\rangle$ will become nearly the same as ε approaches zero. This near optimality along with Hayden and Winter's result has, understandably, largely ceased the study of entan-
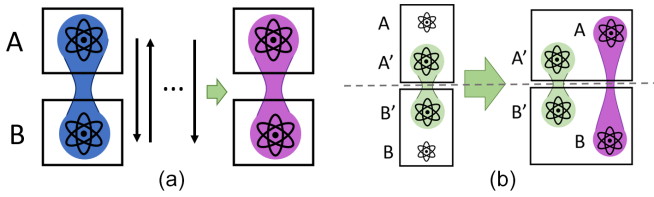
FIG. 1. Conversion of pure quantum states in distant labs. (a) The LOCC model where communication is exchanged. (b) The embezzling of quantum states where an auxiliary entangled state is used. This may be seen as a special case of catalytic conversion.

glement transformations with zero communication because when one needs entanglement transformations without communication, one uses embezzlement [15,16].[1] It is, however, not clear what is the necessary error for embezzlement to become near optimal, which could be relevant in practical settings. Indeed, for any tolerated error, it is easy to find sufficient conditions on pure quantum states to be converted with no catalyst at all (Example 2 of Sec. III). This is an indication that we also do not understand embezzling sufficiently well.

A. Summary of results

The primary aim of this work is to provide tighter lower bounds on the error in pure quantum state entanglement convertibility with zero communication. A high-level comparison of our results to the aforementioned work on this topic is presented in Fig. 2. This depicts a "one-shot resource tradeoff" region that must contain the "true" one-shot resource tradeoff surface for a given pure state conversion. Hayden and Winter's result provides a lower bound on the achievability independent of the amount of shared maximally entangled states, but their result can be too loose when considering zero communication. The result of van Dam and Hayden provides an outer bound on the achievability surface on the face pertaining to LOSE, but their result in fact can be too loose when the error is not sufficiently small. In this work, our results allow one to exactly solve the minimal error in the zero-communication setting and also provide significantly tighter bounds than quantum embezzling for a relevant region on the LOSE face (see Fig. 2).

To formally establish our results, we reduce the class of questions regarding optimal pure quantum state conversion to optimization problems that only concern non-negative vectors. This is because of a bijection between the equivalence classes of pure quantum states under local unitaries, which are defined solely by their Schmidt coefficients, and the probability simplex. We do this by showing the optimal fidelity of pure quantum state transformations with local unitaries is efficiently computable. Of course, in general one would not expect local unitaries to be the optimal strategy and we build on this result to present a nonconvex optimization program over an optimization variable with bounded dimension. An

---

[1] The notable exceptions to this halted topic of research have been the consideration of special embezzling families [17] and the correlated sampling lemma [18], which may be viewed as a variation of embezzling.
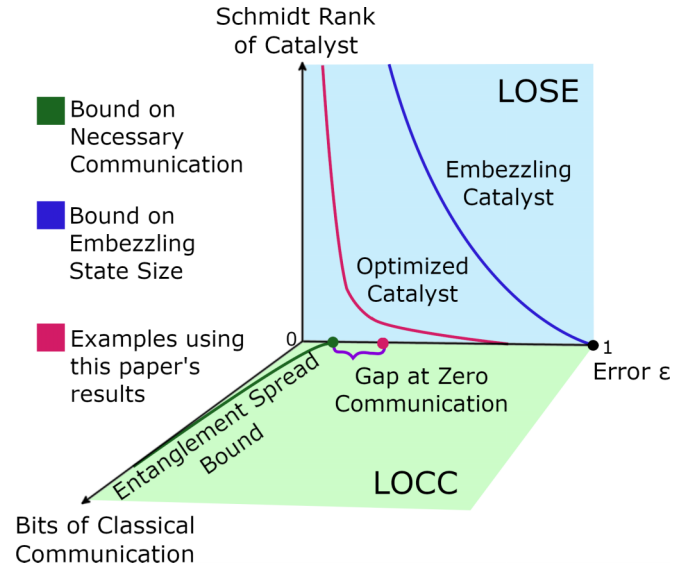

FIG. 2. Comparison of [14] (blue), [19] (green), and this work's results (pink). Reference [19] finds lower bounds on the classical communication necessary to convert one state to another, but in the zero-communication setting the bound is loose (see Examples 1 and 3 for an example of the depicted gap). We find methods for solving this exactly (Sec. IV), which establishes that communication is necessary for larger tolerated errors. Reference [14] establishes a method for pure quantum state transformations with zero communication with massive amounts of entanglement. This result is known to be almost optimal for sufficiently small error. Nonetheless, we find we find the necessary resources can be too strong for a relevant error range, even if ultimately it is optimal (Sec. VI).

immediate corollary of this result is the impossibility of pure quantum state conversions with zero communication for negligible error. We also present efficient computable upper bounds on the achievable error using a semidefinite programming (SDP) relaxation. We also show that in the case where either the seed (i.e., initial) or target state is a two-qubit state, the local unitary strategy is optimal. However, we can show for larger dimensions this is not the case.

Having established general properties in the single-copy case, we move to the multiple-copy case, i.e., where the seed and/or target state is of independent and identically distributed (i.i.d.) form. This is standard in determining the rate of converting one state to another. In particular, we consider dilution and distillation where the seed state or target state, respectively, is many copies of a maximally entangled state and show these are convex optimization programs and may be seen as involving the Ky-Fan norms when extended to the regime where they are not a norm. Lastly, in a sense extending our earlier two-qubit results, we establish that if the target state is an *n*-fold copy of a two-qubit entangled state and the seed state's Schmidt rank is less than the target state, then local unitaries are the optimal strategy.

Finally, given these results, we turn our attention to quantum embezzlement. We begin by noting that the correspondence between Schmidt coefficients and probability distributions means that quantum embezzlement implies a classical equivalent we call randomness embezzlement. We then proceed to use our tools to consider the problem of

approximate catalytic pure quantum state conversion under local unitaries, in effect a generalization of traditional embezzling, and compare it to embezzling. We show in particular that at least in general the optimality of the embezzling states is only for very small errors. Indeed, we show for reasonable tolerable errors, the embezzling state may have a Schmidt rank of many orders of magnitude larger than a state-dependent embezzler. This may have practical relevance and strongly refines our understanding of quantum state transformations under LOSE.

B. Relation to previous work

There has been a great deal of work on embezzlement within communication theory as well as to some degree within quantum thermodynamics. There has also been work on state conversion under LOSR. As such, we briefly state how our work relates to these and other mathematical methodologies. The major technical results of this work focus on the maximum fidelity of conversion between bipartite entangled pure states under local unitaries or local operations and shared randomness. This may be viewed as a variation of optimal conversion distance, which is often described in terms of trace distance (See [20] for a discussion of conversion distance in terms. of trace norm for general resource theories.) This significantly differs from previous works on conversion of bipartite states under LOSR [11,21], which establish equivalent conditions for there being an *exact* conversion from one bipartite state to another under LOSR. As previously stated, we use our technical results to look at embezzling and necessary communication for entangled state transformations. While motivated and focused on the work of van Dam and Hayden [14] and Hayden and Winter [6], there has been further related work. In terms of necessary communication, [22] showed necessary and sufficient conditions for there to exist an exact catalytic transformation between two quantum states under LOCC. In terms of embezzlement, [23] was the first to construct universal embezzling families with respect to zero communication for $m > 2$ parties. Reference [17] studied properties of universal embezzling families with respect to zero communication in more detail. In quantum thermodynamics, there is the notion of "thermal embezzling," which is understood when the states commute with the relevant Hamiltonian, thereby reducing the problem to majorization [24,25]. Reference [24] sketches how quantum embezzlement as done in van Dam and Hayden [14] implies a type of thermal embezzlement that shows the second law of thermodynamics does not hold in an approximate catalytic sense. This formal insight is equivalent to what we call "randomness embezzlement" in Sec. IV. In [25], they can exclude this issue with the second law by imposing physical assumptions. They also construct an optimal universal thermal embezzling family. Finally, we remark that embezzling (with respect to any set of allowed transformations) is a form of approximate catalytic transformation. We refer the reader to recent reviews on catalysis for further information [12,13].

C. Organization of the paper

The rest of the paper is organized as follows. In Secs. II and III we present the necessary notation and background, respectively, to understand the rest of the paper. In Sec. IV, we do the following:

(i) Make explicit the correspondence between pure quantum states under local unitaries LU and the probability simplex and note this implies the existence of a classical variation of embezzlement (Proposition 7).

(ii) Prove our equation for fidelity of state conversion under local unitaries (Theorem 1) and our optimization for fidelity of state conversion under local operations and shared randomness (Theorem 2).

(iii) Establish computable upper bounds on the fidelity of state conversion under LOSR (Theorem 5).

In Sec. V we present the results where the target or seed state is of i.i.d. form. In Sec. VI we discuss embezzlers under local unitaries. In Sec. VII we discuss why our theory does not generalize beyond bipartite pure states.

II. BASIC NOTATION

Our notation largely aligns with standard texts [26,27]. In this paper we consider finite-dimensional quantum systems. Given $n \in \mathbb{N}$, we define $[n] := \{1, \ldots, n\}$. A finite-dimensional Hilbert space will be labeled with a capital roman letter, e.g., A, B, and may be identified by its dimension $d \in \mathbb{N}$, e.g., $A \cong \mathbb{C}^d$. The space of quantum states, or density matrices, with respect to a Hilbert space A, is the space of positive-semidefinite operators with unit trace, i.e., $D(A) := \{\rho \in L(A) : \rho \succeq 0 \text{ and } \operatorname{Tr}(\rho) = 1\}$ where $\succeq$ is the Löwner order and $L(A)$ is the space of endomorphisms. If a quantum state is a joint state over multiple Hilbert spaces, we will use a subscript to specify this, e.g., $\rho_{AB} \in D(A \otimes B)$. A quantum state ρ_A is pure if $\rho_A = |\psi\rangle\langle\psi|$ for a unit vector $|\psi\rangle$, where we are using bra-ket notation. For this previous reason, we generally just specify a pure quantum state by $|\psi\rangle_A$, or ψ if we are considering its density matrix representation. We say a state is classical if it is diagonal in the standard basis, i.e., it is a probability distribution. The space of probability distributions over d elements is denoted $\mathcal{P}(d)$. A quantum channel $\mathcal{E}$ is a (linear) completely positive, trace-preserving map $\mathcal{E} : L(A) \to L(B)$. We denote the set of quantum channels from $L(A) \to L(B)$ as $C(A, B)$.

III. BACKGROUND AND MOTIVATION

Throughout this section we fix $A \cong \mathbb{C}^d$, $B \cong \mathbb{C}^{d'}$ for clarity.

a. Fidelity. The fidelity is a standard measure of similarity between two positive-semidefinite operators $R, S \geqslant 0$:

$$F(R, S) = \|\sqrt{R}\sqrt{S}\|_1^2 = \operatorname{Tr}\left(\sqrt{\sqrt{S}R\sqrt{S}}\right)^2, \qquad (1)$$

where the square root of a positive-semidefinite operator is defined in the standard fashion on its spectral decomposition and $\|\cdot\|_1$ is the Schatten 1-norm. It satisfies various properties that will be relevant for this work which we summarize here. All of these may be verified by direct calculation or by referring to standard texts.

Proposition 1 (Summary of fidelity properties). Let $\rho, \sigma \in D(A)$. The following hold:

(1) $0 \leqslant F(\rho, \sigma) \leqslant 1$ where the upper bound is saturated if and only if $\rho = \sigma$ and the lower bound saturates if and only if their images are orthogonal.

(2) The fidelity is isometrically invariant, i.e., given isometry $V : A \to B$,

$$F(V\rho V^\dagger, V\sigma V^\dagger) = F(\rho, \sigma) .$$

(3) The fidelity satisfies data processing. That is, for any quantum channel $\mathcal{E} \in C(A, B)$,

$$F(\rho, \sigma) \leqslant F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) .$$

(4) If both states are pure,

$$F(|\phi\rangle\langle\phi|, |\psi\rangle\langle\psi|) = |\langle\psi|\phi\rangle|^2 ,$$

and if one state is pure

$$F(|\phi\rangle\langle\phi|, \sigma) = \langle\phi|\sigma|\phi\rangle .$$

(5) If both states are classical, $P, Q \in \mathcal{P}(d)$, then the fidelity reduces to the square of the Bhattacharyya coefficient:

$$F(P, Q) = \left(\sum_{i \in [d]} \sqrt{p(i)q(i)} \right)^2 = \mathrm{BC}(p, q)^2 ,$$

where $p(i) = P(i, i)$ and likewise for Q.

(6) Given pure quantum states with the same eigenbasis and real amplitudes, $|\psi\rangle = \sum_x \sqrt{p(x)}|x\rangle$, $|\phi\rangle = \sum_x \sqrt{q(x)}|x\rangle$, the fidelity reduces to the square of the Bhattacharyya coefficient of the probability distributions defined by the amplitudes

$$F(|\phi\rangle\langle\phi|, |\psi\rangle\langle\psi|) = \mathrm{BC}(p, q)^2 .$$

We also note that in all of these definitions there is a pesky squaring that effectively we do not care about. For this reason we could define the square-root fidelity

$$\sqrt{F}(R, S) := \sqrt{F(R, S)} .$$

Note the square-root fidelity could be viewed as the quantum extension of the Bhattacharyya coefficient.

b. Norms. In defining the fidelity we used the Schatten 1-norm. More generally, there are the Schatten p-norms which for $X \in L(A, B)$ may be defined as $\|X\|_p := \|\sigma(X)\|_p$ where $\sigma(X)$ is the ordered vector of singular values of X, $\sigma_1(X) \geqslant \sigma_2(X) \geqslant \cdots \geqslant \sigma_{\mathrm{rank}(X)}(X)$ and it is being evaluated under the L_p-norm where $p \geqslant 1$. The infinity norm, ∞-norm, is $\lim_{p \to \infty} \|X\|_p = \|X\|_\infty = \max_i \sigma_i(X)$. The infinity norm was generalized to the Ky-Fan k-norms $\|X\|_{(k)} := \sum \sigma_i(X)$ for $1 \leqslant k \leqslant \min\{d, d'\}$. The Ky-Fan norms have relevance in measuring entanglement [28]. A generalization of the Ky-Fan and Schatten norms together is given by the (k, p)-norms [29]

$$\|X\|_{(k,p)} := \left(\sum_{i \in [k]} \sigma_i(X)^p \right)^{1/p} , \tag{2}$$

which also have use in measuring entanglement of pure quantum states [30]. Much like is common to do for the Schatten p-norms, we can extend the (k, p)-norms to $p > 0$ with the caveat they will not be norms as they will not in general satisfy subadditivity (the triangle inequality) for $p \in [0, 1)$.

c. Entanglement theory. A bipartite positive operator ρ_{AB} is separable if there exists $n \in \mathbb{N}$, $p \in \mathcal{P}(n)$, $\{\sigma_A^i\}_{i \in [n]} \subset D(A)$, and $\{\tau_B^i\}_{i \in [n]}$ such that

$$\rho_{AB} = \sum_{i \in [n]} p(i)\sigma_A^i \otimes \tau_B^i .$$

Otherwise, the state is entangled. As a pure quantum state $|\psi\rangle\langle\psi|_{AB}$ is defined by a unit vector, this reduces to a pure quantum state is separable, referred to product in this setting, if and only if there exists $|\phi\rangle_A, |\varphi\rangle_B$ such that $|\psi\rangle = |\phi\rangle_A \otimes |\varphi\rangle_B$. While this is sufficient for determining if a bipartite pure state is entangled, there is also a notion of "how" entangled a state is in terms of Schmidt rank. Every bipartite pure state $|\psi\rangle_{AB}$ admits a unique (up to reordering) decomposition of the form

$$|\psi\rangle_{AB} = \sum_{i \in [k]} \sqrt{p(i)}|u_i\rangle_A \otimes |v_i\rangle_B, \tag{3}$$

where $k = \max\{d, d'\}$, $p \in \mathcal{P}(k)$, and $\{|u_i\rangle\}_{i \in [k]}$, $\{|v_i\rangle\}_{i \in [k]}$ are orthonormal bases of A and B, respectively. We note that for nonunit vectors, such a decomposition still exists, it is just $p \in \mathbb{R}_{\geqslant 0}$, i.e., is a non-negative vector rather than a probability distribution.

The $\sqrt{p(i)} > 0$ terms are referred to as the Schmidt coefficients of the pure quantum state. The Schmidt rank of $|\psi\rangle_{AB}$, $\mathrm{SR}(|\psi\rangle) = \mathrm{supp}(p)$, i.e., the number of Schmidt coefficients. This may be viewed as a measure of entanglement in the sense that the Schmidt rank of a product state is 1 and the maximally entangled state $|\Phi^+\rangle_{\mathbb{C}^d\mathbb{C}^d} = \frac{1}{\sqrt{d}} \sum_i |i\rangle_{\mathbb{C}^d}|i\rangle_{\mathbb{C}^d}$ has Schmidt rank d. We define the set $\mathrm{SR}(d) := \{|\psi\rangle : \mathrm{SR}(|\psi\rangle) \leqslant d\}$, where we note this set is independent of the dimension the state is embedded in.

Lastly we note a particularly nice property of pure states, known as Uhlmann's theorem.

Lemma 1 (Uhlmann's theorem). Given $R, T \in \mathrm{Pos}(A)$ and (possibly unnormalized) $|\psi\rangle \in A \otimes B$ such that $\mathrm{Tr}_B(\psi) = R$, then

$$F(R, T) = \max\{|\langle\psi|\phi\rangle|^2 : |\phi\rangle \in A \otimes B, \ \mathrm{Tr}_B(\phi) = T\} .$$

d. No-go theorems, embezzling, and motivation. With the established background, we now present the previous results related to zero-communication pure state transformations which we will discuss our results in relation to. The first is a lower bound on the number of qubits or classical bits necessary to convert between pure quantum states [6].

Proposition 2. (Reference [6, Theorem 8]). Consider a state transformation via an arbitrary channel $\mathcal{E} \in C(A \otimes B, A \otimes B)$ from seed state $|\phi\rangle_{AB}$ to target state $|\psi\rangle_{AB}$ such that $F(\mathcal{E}(\phi), \psi) \geqslant 1 - \varepsilon$ where the channel $\mathcal{E}$ is implemented using local operations and quantum communication. Then, independent of any amount of entanglement assistance, for $\delta = \sqrt[8]{\varepsilon}$, in the implementation of $\mathcal{E}$, q qubits must have been exchanged where

$$q \geqslant \frac{1}{2}[\Delta_\delta(\mathrm{Tr}_B(|\psi\rangle\langle\psi|)) - \Delta_0(\mathrm{Tr}_B(|\phi\rangle\langle\phi|))] + \log_2(1 - \delta), \tag{4}$$

where

$$\exp[\Delta_\varepsilon(P)] = \min \ \mathrm{rank}(\widetilde{P})\lambda_{\max}(\widetilde{P})$$
$$\mathrm{s.t.} \mathrm{Tr}(\widetilde{P}) \geqslant 1 - \varepsilon,$$
$$\widetilde{P} = \Pi P \Pi,$$
$$[P, \Pi] = 0,$$
$$\Pi^2 = \Pi .$$

Moreover, the bound given in (4) holds for a necessary amount of classical communication by multiplying the right-hand side by two.

While the above proposition is very powerful and implies two states with different Schmidt decompositions cannot be perfectly converted with zero communication, it is not sufficient in every scenario. In particular, the following example shows that in certain cases Proposition 2 cannot eliminate *any* state from being able to be converted to a given target state with relatively high fidelities.

Example 1 (On the necessity of communication). Up to local unitaries, let the target state be $|\psi\rangle = \sqrt{0.54}|00\rangle + \sqrt{0.02}|11\rangle + \sqrt{0.44}|22\rangle$. Assume we are interested in a state transformation $\mathcal{E}$ such that $F(\mathcal{E}(\phi), \psi) = 0.99$, where $|\phi\rangle$ is the seed state. Then $\varepsilon = 0.01$, so $\delta > 0.56$. Note $\mathrm{Tr}_B(|\psi\rangle\langle\psi|) = 0.54|0\rangle\langle0| + 0.02|1\rangle\langle1| + 0.44|2\rangle\langle2|$. Then $\Delta_\delta(\mathrm{Tr}_B(|\psi\rangle\langle\psi|)) = \log_2(|1| \times 0.44) < -1.18$, by removing the 0.02 and 0.54 eigenvalues. It may be shown [6] that $\Delta_0(\mathrm{Tr}_B(|\psi\rangle\langle\psi|)) \geqslant 0$ and clearly $\log_2(1 - \delta) < 0$. It follows that in this setting the right-hand side of (4) is negative. Therefore, we have no proof from this bound that any transformation for any seed state which achieves this relatively high fidelity of 99% requires any communication.

While the above example shows there are reasonably small tolerated errors ε where Proposition 2 is not helpful, when the tolerated error is sufficiently small, it will imply the need for communication. This sort of structure for sufficiently small ε also appears when considering quantum embezzlement [14], which may be seen as a solution to Proposition 2 implying communication is necessary. Quantum embezzlement in effect shows one can make pure state transformations with zero communication to any nonzero error if they have the right sufficiently large embezzling state.

Proposition 3. (Reference [14]). Consider the family of embezzler states $|\mu(n)\rangle_{A'B'} = \frac{1}{\sqrt{H_n}}\sum_{j=1}^{n}\frac{1}{\sqrt{j}}|j\rangle_{A'}|j\rangle_{B'}$ where $H_n := \sum_{i=1}^{n} n^{-1}$ is the harmonic number. For any $\varepsilon > 0$ and target bipartite pure quantum state $|\psi\rangle_{AB}$ with Schmidt rank m, for $n > m^{1/\varepsilon}$ there exist unitaries $U_{AA'}, W_{BB'}$ such that

$$F[U_{AA'} \otimes W_{BB'}(|\mu(n)\rangle_{A'B'}|0\rangle_A|0\rangle_B],$$
$$|\mu(n)\rangle_{A'B'} \otimes |\psi\rangle_{AB}) \geqslant 1 - \varepsilon.$$

Moreover, U, W are in effect permutations on the joint Schmidt bases.

One can see quantum embezzlement implies a way to convert one pure state to another to non-zero error by picking a large enough embezzler and then first "embezzling out" the original state (uncomputing $|\phi\rangle$ to $|0\rangle|0\rangle$ via embezzling) and then "embezzling in" the target state $|\psi\rangle$.

What is perhaps most remarkable about the above approach is that it was shown in the original work that even if we allow

LOCC and a state-dependent embezzler $|\zeta\rangle$, the amount of entanglement n in $|\zeta\rangle$ must scale proportionally to $\frac{1}{\log_2(n)}$ as the error ε becomes close to zero, but the universal embezzling family $|\mu(n)\rangle$ scales the same way in the entanglement n. That is, as ε approaches zero, embezzling using van Dam and Hayden's universal embezzling family is effectively optimal. However, just as with the discussion pertaining to Proposition 2, it is clear embezzling is not necessary for reasonable error levels in general. In fact, we show in the following example that for any nonzero error there exist states which can be converted without any catalyst or embezzler.

Example 2 (On the necessity of embezzling). As noted, as $\varepsilon \to 0$, embezzling is necessary. However, it is not in general clear at what point embezzling becomes necessary. This can be seen as follows. Consider $\varepsilon \in (0, 1)$ and two probability distributions $p, q \in \mathcal{P}(m)$ such that the $\mathrm{BC}(p, q)^2 \geqslant 1 - \varepsilon$. Define the seed state as $|\phi\rangle = \sum_{i\in[m]}\sqrt{p(i)}|i\rangle_A|i\rangle_B$ and the target state as $|\psi\rangle = \sum_{i\in[m]}\sqrt{q(i)}|i\rangle_A|i\rangle_B$. Then we have

$$F(|\phi\rangle\langle\phi|, |\psi\rangle\langle\psi|) = \mathrm{BC}(p, q)^2 \geqslant 1 - \varepsilon ,$$

where we have used item 5 of Proposition 1. Therefore, given $|\phi\rangle$, it requires *no* communication or entanglement to generate $|\psi\rangle$ to error ε. In fact, as we show later (Proposition 4), this will be true for converting the set of states with Schmidt coefficients defined via p to the set of states with Schmidt coefficients defined via q in general.

Given these two examples, we see that while these results give strong characterizations of pure state transformations with zero communication, neither the need for communication by Proposition 2 nor the optimality of Proposition 3 when the error tends to zero give us a full understanding of this setting. It would therefore be of value to better understand this task, and this is what the rest of this work addresses.

IV. SINGLE-COPY PURE STATE CONVERSION WITH ZERO COMMUNICATION

Our primary goal of this section is to determine the minimal error of conversion between pure quantum states with zero communication, which would resolve the gap presented in Example 1. To establish the minimal error of conversion, we will use the correspondence between the probability simplex and Schmidt coefficients under local unitaries (LU), which we establish in the following subsection. We also note that this implies the existence of a classical equivalent of embezzling, which we call randomness embezzling (Theorem 5). This correspondence motivates the idea that the optimal fidelity of pure state conversion under local unitaries is simply reordering the Schmidt coefficients, which we in fact prove (Theorem 1). We then use the local unitary result to establish a bounded but nonlinear optimization program that determines the optimal achievable fidelity under conversion via local operations and shared randomness (LOSR), which does not require shared randomness (Theorem 2). We end the section by discussing the relationship between the LU and LOSR strategies and introducing an SDP relaxation for efficiently establishing upper bounds on the achievable fidelity of pure state conversions under LOSR.

A. Correspondence under local unitaries between Schmidt coefficients and the probability simplex

In this subsection we establish the bijection between Schmidt coefficients, which define the equivalence classes of bipartite pure quantum states under local unitaries, and the probability simplex. One reason for this is because the rest of the results of this work might be best seen as verifying that in the zero-communication setting this correspondence is all that matters. Indeed, we will see this in the subsequent subsections which show that the minimal fidelity error of pure state transformations under zero communication will always be functions of only the Schmidt coefficients.

Definition 1. We define the the ordered probability simplex $\mathcal{P}^{\downarrow}(d)$ as $p \in \mathcal{P}^{\downarrow}$ if $p(i) \geqslant p(i+1)$ for all $i \in [d-1]$.

Proposition 4. Up to local unitaries, any pure quantum state is of the form

$$|\psi\rangle_{AB} = \sum_{i \in [k]} \sqrt{p^{\downarrow}(i)} |i\rangle_A \otimes |i\rangle_B,$$

where $p^{\downarrow}(i) \geqslant p^{\downarrow}(i+1)$ for all $i \in [k-1]$, $k = \max\{d, d'\}$, $p^{\downarrow} \in \mathcal{P}^{\downarrow}(k)$, and $\{|i\rangle\}$ is the computational basis in both cases. In other words, there exist both equivalence classes on pure quantum states under local unitary operations in terms of Schmidt coefficients and ordered Schmidt coefficients.

Proof. Consider $|\psi\rangle_{AB} = \sum_{j \in [k]} \sqrt{p'(j)} |u_j\rangle \otimes |v_j\rangle$ as decomposed in (3). Now fix the permutation π on $[k]$ such that $p'[\pi^{-1}(i)] \geqslant p'[\pi^{-1}(i+1)]$ for all $i \in [k-1]$, i.e., π relabels p' so that it is decreasing. Define the unitaries $U_A = \sum_{j \in [k]} |\pi(j)\rangle\langle u_j|$, $W_B = \sum_{j \in [k]} |\pi(j)\rangle\langle v_j|$, which may be verified to be unitaries by direct calculation. Then $(U_A \otimes W_B)|\psi\rangle_{AB}$ will be of the form given in the proposition statement. Finally, we could make this argument for any pure quantum state without ordering the Schmidt coefficients to get one set of equivalence classes. As such, under local unitaries, we can define equivalence classes of pure quantum states in terms of ordered or nonordered Schmidt coefficients. This completes the proof. ∎

Definition 2. The space of (representatives of the equivalence class of) ordered Schmidt coefficient pure quantum states with Schmidt rank bounded by d is given by $\mathrm{SR}^{\downarrow}(d)$. That is, if $|\psi\rangle \in \mathrm{SR}^{\downarrow}(d)$, then $|\psi\rangle = \sum_{i \in [d]} \sqrt{p^{\downarrow}(i)} |u_i\rangle |i\rangle |i\rangle$ where $p^{\downarrow} \in \mathcal{P}^{\downarrow}(d)$.

We can use the previous proposition to relate the (ordered) probability simplex over d elements to to the equivalence classes of (ordered) Schmidt decompositions with Schmidt rank bounded by d. This will make use of the vec mapping.

Definition 3. Given the space of linear operators $\mathrm{L}(A, B)$ where $A \cong \mathbb{C}^d$, $B \cong \mathbb{C}^{d'}$, the vec mapping $\mathrm{vec} : \mathrm{L}(A \otimes B) \to A \otimes B$ is defined by $\mathrm{vec}(|i\rangle\langle j|) = |j\rangle \otimes |i\rangle$ where $\{|i\rangle\}_{i \in [d]}$ and $\{|j\rangle\}_{j \in [d']}$ are the computational bases for A and B, respectively.

This choice of definition for the vec mapping satisfies the identity

$$\left(X_1^T \otimes X_0\right) \mathrm{vec}(Y) = \mathrm{vec}(X_0 Y X_1), \tag{5}$$

where $X_0 \in \mathrm{L}(A_0, B_0)$, $X_1 \in \mathrm{L}(A_1, B_1)$, and $Y \in \mathrm{L}(B_1, B_0)$.

Proposition 5. Consider the functions $\mathrm{vec}(\sqrt{\cdot}) : \mathrm{L}(\mathbb{C}^d) \to \mathbb{C}^d \otimes \mathbb{C}^d$ and $\mathrm{vec}^{-1}(\cdot^{\odot 2}) : \mathbb{C}^d \otimes \mathbb{C}^d \to \mathrm{L}(\mathbb{C}^d)$ where $\cdot^{\odot 2}$ is the entrywise square of a vector. These functions define a

bijection between $\mathcal{P}(d)$ [respectively $\mathcal{P}^{\downarrow}(d)$] and the space of equivalence classes of Schmidt decompositions under local unitaries with Schmidt rank bounded by d [respectively the space $\mathrm{SR}^{\downarrow}(d)$].

Proof. We prove it via direct calculation for $\mathcal{P}(d)$ and the space of Schmidt decompositions. The proof in the other case works the same. Let $C \cong \mathbb{C}^d$. First, consider $p \in \mathcal{P}(d)$ which we write in its density matrix form, e.g., $P = \sum_{i \in [d]} p(i)|i\rangle\langle i|$. Then

$$\mathrm{vec}(\sqrt{P}) = \mathrm{vec}\left(\sum_{i \in [d]} \sqrt{p(i)}|i\rangle\langle i|\right)$$
$$= \sum_{i \in [d]} \sqrt{p(i)}|i\rangle_C \otimes |i\rangle_{C'},$$

which is in the specified equivalence class by applying isometries that take the computational bases from C, C' to A, B. In the other direction, take the Schmidt decomposition in the purified basis $|\psi\rangle_{AB} = \sum_{i \in [d]} \sqrt{q(i)}|i\rangle_A \otimes |i\rangle_B$. We can convert the A space to C via the channel

$$\mathcal{F}_{A \to C}(\cdot) := V^{\dagger} \cdot V + (\mathbb{1} - V^{\dagger}V) \cdot (\mathbb{1} - V^{\dagger}V),$$

where $V = \sum_{i \in [d]} |i\rangle_A\langle i|_C$ is the isometry that takes the C space to the A space as $|A| \geqslant |C|$ by assumption. The same type of conversion holds for the B and C'. Therefore, we have (up to equivalences) $|\psi\rangle_{AB} = \sum_{i \in [d]} \sqrt{q(i)}|i\rangle_C|i\rangle_{C'}$. Then,

$$\mathrm{vec}^{-1}(|\psi\rangle^{\cdot 2}) = \mathrm{vec}^{-1}\left(\sum_{i \in [d]} q(i)|i\rangle_C|i\rangle_{C'}\right)$$
$$= \sum_{i \in [d]} q(i)|i\rangle\langle i|_C,$$

where in the last line we used that $C' \cong C$ so that $\mathrm{L}(C, C') \cong \mathrm{L}(C)$. This completes the proof. ∎

The reason this is useful is it draws equivalence between the equivalence classes of entangled states in terms of Schmidt coefficients and probability distributions under fidelity.

Proposition 6. Consider $|\phi\rangle = \sum_{i \in [d]} \sqrt{p(i)}|i\rangle_A|i\rangle_B$, $|\psi\rangle = \sum_{i \in [d]} \sqrt{q(i)}|i\rangle_A|i\rangle_B$. Then $F(|\phi\rangle\langle\phi|, |\psi\rangle\langle\psi|) = \mathrm{BC}(p, q)^2$.

Proof. First note $V : |i\rangle_A \to |i\rangle_A|i\rangle_B$ is an isometry. Define $|\phi'\rangle := \sum_i \sqrt{p(i)}|i\rangle_A$ and similarly $|\psi'\rangle$. Note $V|\phi'\rangle = |\phi\rangle$ and $V|\psi'\rangle = |\psi\rangle$. Thus, we have

$$F(|\phi\rangle\langle\phi|, |\psi\rangle\langle\psi|) = F(V|\phi'\rangle\langle\phi'|V^{\dagger}, V|\psi'\rangle\langle\psi'|V^{\dagger})$$
$$= F(|\psi'\rangle\langle\psi'|, |\phi'\rangle\langle\phi'|)$$
$$= \mathrm{BC}(p, q)^2,$$

where the first equality is our observation, the second is isometric equivalence (item 2 of Proposition 1), and the third is using item 6 of Proposition 1 where we note $|\psi'\rangle, |\phi'\rangle$ are of the form given in that item. This completes the proof. ∎

Randomness embezzling. Before moving forward, we note that independent of the focus of this work, this equivalence between Schmidt coefficients and the probability simplex means that the proof of quantum embezzlement also proves the existence of a classical version. Specifically, the proof of quantum embezzlement [14] only bounds the fidelity between
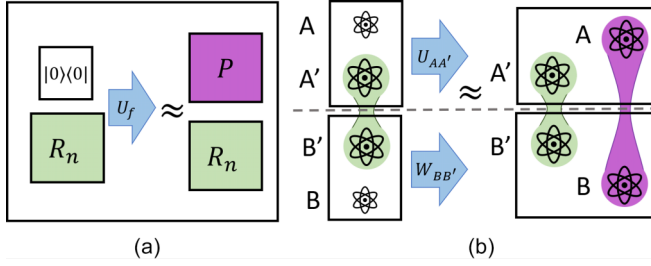
FIG. 3. Comparison between embezzlement of classical distributions and quantum states. (a) The embezzlement of classical distributions happens within one laboratory and a local permutation of the joint computational basis. (b) The embezzling of quantum states happens across two laboratories where each party applies the permutation of the joint computational basis on their local halves. Note in both cases this is done in an approximate fashion as denoted by the $\approx$ symbol in the diagrams.

the Schmidt coefficients of the embezzling state and the target state tensored with the embezzling state (under some permutation). This allows them to bound the fidelity by reducing it to the Bhattacharyya coefficient of the Schmidt coefficients:

$$F(|\psi\rangle, |\phi\rangle) = |\langle \psi, \phi \rangle|^2 = |\langle \sqrt{P}, \sqrt{Q} \rangle|^2$$
$$= \left(\sum_i \sqrt{p(i)q(i)} \right)^2$$
$$= \mathrm{BC}(p, q)^2 .$$

This argument follows the same form as the previous few propositions. This allows us to ultimately conclude the same proof bounds a classical equivalent of embezzling (Proposition 7) which is depicted in Fig. 3 for comparison with quantum embezzlement. This is shown in further formal detail in Appendix A.

Moreover, we note this above idea has been previously explored in the context of quantum thermodynamics. Specifically, this exact idea was sketched in [24] where it was used to show that any process can be done cyclically when an approximate error condition is permitted and there are no restrictions on the embezzling state. Moreover, Ref. [25] explored this issue in further detail and in particular showed thermal embezzling no longer violates the second law in an approximate sense when there are physical restrictions on the embezzling state. As we did not present the proof for embezzlement of quantum states and the proof is omitted in [24], we present the proof of embezzlement of probability distributions in full for clarity in Appendix A.

Proposition 7. (See also [24,25]). For any $\varepsilon > 0$ and target probability distribution $P \in \mathcal{P}(m)$, the embezzling distribution $R_n := \frac{1}{H_n} \sum_{j=1}^{n} \frac{1}{j} |j\rangle\langle j|$ is such that for $n > m^{1/\varepsilon}$ there exists a unitary representation of a basis relabeling U_f of the joint distribution such that

$$F(U_f(R_n \otimes |0\rangle\langle 0|)U_f^\dagger, R_n \otimes P) \geqslant 1 - \varepsilon.$$

We note the major difference between randomness and quantum embezzlement is the role of locality. In the classical case there is a single party and the distribution is not bipartite, both of which remove the notion of locality. These differences

are nontrivial: one cannot construct a nonlocal classical equivalent of embezzling that at the same time demands that the embezzler remains decoupled as in Proposition 3, and one cannot find a quantum equivalent of the nonlocal classical variation that one can implement as follows from Proposition 2. As it is not central to the rest of this work, we provide an extended discussion of this nuance for the interested reader in Appendix A after the proof of Theorem 5.

B. Pure state conversion under local unitaries

Having established the relationship between the equivalence classes of pure states in terms of Schmidt coefficients and the probability simplex, we now show the optimal strategy for converting one pure quantum state to another under local unitaries is simply relabeling the Schmidt basis so the ordering of the Schmidt coeffficients is the same. This is not necessarily surprising. It is not clear what more one could do, and indeed this is the strategy that is used to implement quantum embezzlement [14].

Lemma 2. Let $R \in \mathrm{Pos}(\mathbb{C}^d)$, $T \in \mathrm{Pos}(\mathbb{C}^{d'})$. Then

$$\max_U F(P, UQU^\dagger) = F(\widetilde{R}^\downarrow, \widetilde{T}^\downarrow) ,$$

where $R^\downarrow = \sum_i \nu_i(R)|i\rangle\langle i|$, $\nu_1(R) \geqslant \nu_2(R) \geqslant \cdots$ are the (decreasing) ordered eigenvalues of R, and likewise for $T^\downarrow$. In other words, the fidelity between R and T maximized over unitaries is equal to the fidelity of their ordered eigenvalues.

Proof. This proof is a combination of the definition of fidelity and a corollary of von Neumann's trace theorem. A similar identity was established in [31]:

$$\max_U F(R, UTU^*) = \max_U \|\sqrt{R}\sqrt{UTU^*}\|_1^2$$
$$= \max_U \|\sqrt{R}U\sqrt{T}U^*\|_1^2$$
$$= \max_U (\max_W |\mathrm{Tr}[W\sqrt{R}U\sqrt{T}U^*]|)^2$$
$$= \max_{U,W} |\mathrm{Tr}[W\sqrt{R}U\sqrt{T}]|^2$$
$$= \left(\sum_{i \in [q]} \sigma_i(R)\sigma_i(T) \right)^2$$
$$= F(\widetilde{R}^\downarrow, \widetilde{T}^\downarrow).$$

The first equality is (1). The second equality is because $\sqrt{UTU^*} = \sum_i \sqrt{\lambda_i(T)}|\psi_i\rangle\langle\psi_i| = U(\sum_i \sqrt{\lambda_i(T)}|\phi_i\rangle\langle\phi_i|)U^* = U\sqrt{T}U^*$ by defining $|\psi_i\rangle := U|\phi_i\rangle$ where $\{|\phi_i\rangle\}_i$ is the eigenbasis of T. The third equality is a well-known variational form of the 1-norm [27]. The fourth is using cyclicity of trace, redefining $U^*W \to W$, and pulling out the maximization. The fifth is [32, Corollary 7.4.1.3], and the final equality is by definition of the operators (which are defined in the same basis). ∎

We now can use the above lemma to establish the pure state property we are actually interested in. For notational simplicity, we define the following notation:

$$F_{\mathrm{LU}}(\rho, \sigma) := \max_{U,V} F(\rho, (U \otimes V)(\sigma)), \qquad (6)$$

which is without loss of generality unitaries as we can just trivially embed the states ρ, σ so that they both are defined on the same local spaces. That is, in general one would optimize over isometries, but by treating the states embedded into the same local spaces already, we can focus on unitaries. To see this, consider ρ_{AB} and $\sigma_{A'B'}$ such that $|A| =: d_A > d_{A'} := |A'|$. We can then embed the local space A' into A by taking any bases of A and A', respectively, e.g., $\{|\chi_i\rangle\}_{i \in [d_A]}$ and $\{|\zeta_j\rangle\}_{j \in [d_{A'}]}$, and defining the isometry $W_{A' \to A} := \sum_{j \in [d_{A'}]} |\chi_i\rangle\langle\zeta_i|$. We can then consider $W \sigma W^*$ instead of σ directly. The same idea works for embedding B' into B when $|B'| \leqslant |B|$ and we can always treat ρ as in the higher-dimensional space (again, by embedding it). All of this is because it does not affect the eigenvalues of the respective or Schmidt coefficients of the relevant states. Finally, the choice of isometries that we do the embedding for does not matter as the U, V we optimize over can take any choice of the local isometries on the σ state to any others.

Theorem 1. Let $|\psi\rangle, |\phi\rangle \in A \otimes B$ be (possibly unnormalized) vectors with (possibly unnormalized) Schmidt coefficients $r_1 \geqslant r_2 \geqslant \cdots, t_1 \geqslant t_2 \geqslant \cdots$, respectively. Then,

$$\max_{U,V} F(|\psi\rangle, U \otimes V |\phi\rangle) = F(R^\downarrow, T^\downarrow), \qquad (7)$$

where $R^\downarrow = \sum_i r_i |i\rangle\langle i|$, $T^\downarrow = \sum_i t_i |i\rangle\langle i|$. In particular, if we treat these as quantum states, this gives the optimal fidelity of converting $|\phi\rangle$ to $|\psi\rangle$ under local unitaries.

Proof. Up to local unitaries, $|\psi\rangle = \sum_i \sqrt{r_i} |i\rangle |i\rangle$. Therefore, without loss of generality, that can be taken as our target state by allowing free local unitaries on the seed state. We can take the seed state to be of the form $|\phi\rangle = \sum_i \sqrt{t_i} |i\rangle |i\rangle$ by the same argument. Then by assumption, we are interested in $\max_{U,V} F(|\psi\rangle, (U \otimes V)|\phi\rangle)$ with the specified forms. Note

$$\mathrm{Tr}_B[(U \otimes V)|\phi\rangle\langle\phi|(U \otimes V)^\dagger]$$

$$= \sum_{i,i'} \sqrt{t_i t_{i'}} U |i\rangle\langle i'| U^\dagger \mathrm{Tr}(V|i\rangle\langle i'|V^\dagger)$$

$$= \sum_i t_i U |i\rangle\langle i| U^\dagger =: UQU^\dagger.$$

Now for any unitary U we define the following purification:

$$|w^{|U}\rangle := \mathrm{vec}(\sqrt{UQU^\dagger})$$

$$= (\overline{U} \otimes U) \mathrm{vec}(\sqrt{Q}) = (\overline{U} \otimes U)|\phi\rangle,$$

where we have used $\sqrt{UQU^\dagger} = U\sqrt{Q}U^\dagger$ and the vec map identity (5). Now we have

$$F(R^\downarrow, UQU^\dagger) = \max_{|w'\rangle} F(|\psi\rangle, |w'\rangle)$$

$$= \max_V F(|\psi\rangle, (\mathbb{1} \otimes V)|w^{|U}\rangle) \qquad (8)$$

$$= \max_V F(\psi, (\overline{U} \otimes VU)|\phi\rangle),$$

where the first equality is by Uhlmann's theorem (Lemma 1), the second is because all purifications of a given operator are unitarily equivalent on the purifying space [27], so there exists a V such that $(\mathbb{1} \otimes V)|w^{|U}\rangle = |w'\rangle$. The final line is just expanding the definition of $|w^{|U}\rangle$.

It follows

$$\max_{W,V} F(|\psi\rangle, (W \otimes V)|\phi\rangle) = \max_{\overline{U},V'} F(|\psi\rangle, (\overline{U} \otimes V'U)|\phi\rangle)$$

$$= \max_{\overline{U},V'} F(|\psi\rangle, (\mathbb{1} \otimes V)|w^{|U}\rangle)$$

$$= \max_U F(R^\downarrow, UQU^\dagger)$$

$$= F(R^\downarrow, T^\downarrow),$$

where the first equality is because unitaries are closed under multiplication and the optimizations are independent, the second and third are both by (8) for clarity, the third is because unitaries are closed under conjugation, and then the final equality is by applying Lemma 2. This completes the proof. ■

This means under local unitaries, it is efficient to compute the optimal fidelity and that in fact the optimal strategy is simply Alice and Bob reordering the basis so that the Schmidt coefficients are in the same relative ordering. It also follows from item 1 of Proposition 1 that unless all the Schmidt coefficients are equal, the fidelity cannot be one under local unitary strategies.

C. Pure state conversions under local operations and shared randomness

While the previous section is nice in that it finds an efficient way of calculating the optimal conversion strategy under local unitaries, it would be natural to ask if local operations can do better than local unitaries as it is a much more general class of operations. In fact, we can see that it must do better in some cases in a trivial manner. Consider the target state $|\psi\rangle$ and the seed state $|\phi\rangle = |\psi\rangle \otimes |\zeta\rangle$ where $|\zeta\rangle$ is not product. Under local unitaries this transformation is not possible to arbitrary precision because of $|\zeta\rangle$, but of course in reality the parties could trace out whichever portion(s) of $|\zeta\rangle$ they hold. Thus, we need a theory of transformations under local operations.

Note that this trivial example we have given would not be resolved by local mixed unitary strategies, i.e., strategies where each party varies their choice of unitary according to some local randomness. Indeed, we begin by noting that local mixed unitary strategies cannot ever outperform local unitary strategies.

Corollary 1. Let $|\psi\rangle$ be the target state and $|\phi\rangle$ be the seed state and only optimize over Alice and Bob using mixed unitary channels. Then the optimal is the same as in Theorem 1.

Proof. Letting $\mathcal{E}_U, \mathcal{F}_W$ be local mixed unitary maps,

$$\max_{\mathcal{E}_U, \mathcal{F}_W} F(\psi, (\mathcal{E}_U \otimes \mathcal{F}_W)(\phi))$$

$$= \langle\psi|(\mathcal{E}_U \otimes \mathcal{F}_W)(\phi)|\psi\rangle$$

$$= \int_{U,W} \langle\psi|(U \otimes W)(\phi)|\psi\rangle dU \, dW$$

$$\leqslant \int_{U,W} \max_{U,W} \langle\psi|(U \otimes W)(\phi)|\psi\rangle$$

$$= \max_{U,W} \langle\psi|(U \otimes W)(\phi)|\psi\rangle$$

$$= F(P^\downarrow, Q^\downarrow),$$

where the first equality is by item 4 of Proposition 1, the second is letting the mixed unitary map be for any probability measures dU, dW over the unitary group. The inequality is because the inner product is real and so it is lower bounded by the maximum. The second to last equality is by linearity, and the final equality is by Theorem 1. Noting that a specific choice of local unitaries is a special case of mixed unitary channels completes the proof. ∎

The above tells us that we must escape the use of unitaries to improve our bounds. Note, however, that in general the only maps that preserve pure quantum states are isometries, and our results so far have been in terms of pure quantum states, so we need to maintain this structure to build on them. For this reason, the following proof will make use of the isometric representation of quantum channels.

For notational simplicity, we define the optimal fidelity of conversion under local operations and shared randomness (LOSR) fidelity

$$F_{\text{LOSR}}(\rho, \sigma) := \max_{\mu, \mathcal{E}_\lambda, \mathcal{F}_\lambda} F\left(\rho, \int (\mathcal{E}_\lambda \otimes \mathcal{F}_\lambda)(\sigma) d\mu(\lambda)\right),$$

where μ is a probability measure over an index set for sets of local channels $\{\mathcal{E}_\lambda\}$ and $\{\mathcal{F}_\lambda\}$. Similarly, we can define optimal fidelity of conversion under local operations (LO) as

$$F_{\text{LO}}(\rho, \sigma) := \max_{\mathcal{E}, \mathcal{F}} F(\rho, \mathcal{E} \otimes \mathcal{F})(\sigma)).$$

With these defined, we prove the following.

Theorem 2. Let $|\psi\rangle, |\phi\rangle \in A \otimes B$ be (possibly unnormalized) vectors with (possibly unnormalized) Schmidt coefficients $r_1 \geqslant r_2 \geqslant \ldots$ and $t_1 \geqslant t_2 \geqslant \ldots$, respectively. Then,

$$F_{\text{LOSR}}(|\psi\rangle, |\phi\rangle) = F_{\text{LO}}(|\psi\rangle, |\phi\rangle)$$
$$= \max_{P' \in \mathcal{P}(\Sigma)} F((R \otimes P')^\downarrow, T_{\text{embed}}^\downarrow), \quad (9)$$

where the finite alphabet Σ satisfies $|\Sigma| \leqslant [\text{SR}(|\phi\rangle)\text{SR}(|\psi\rangle)]$, $R = \sum_i r_i |i\rangle\langle i|$ and similarly for T_{embed} which is the distribution T embedded into the joint probability simplex over the finite alphabet indexing R and Σ. In particular, if we treat $|\psi\rangle, |\phi\rangle$ as quantum states, this gives the optimal fidelity of converting $|\phi\rangle$ to $|\psi\rangle$ under local operations and shared randomness.

Proof. The first equivalence follows similarly to the mixed unitary case. Clearly the class of LOSR strategies is more general than the class of LO strategies, so we just need to show LOSR is only as strong as LO here:

$$F_{\text{LOSR}}(\phi, \psi) = F\left(\psi, \int (\mathcal{E}_\lambda \otimes \mathcal{F}_\lambda)(\phi) d\mu(\lambda)\right)$$
$$= \int \langle \psi | (\mathcal{E}_\lambda \otimes \mathcal{F}_\lambda)(\phi) | \psi \rangle d\mu(\lambda)$$
$$\leqslant \int \max_{\mathcal{E}, \mathcal{F}} [\langle \psi | (\mathcal{E} \otimes \mathcal{F})(\phi) | \psi \rangle] d\mu(\lambda)$$
$$= \max_{\mathcal{E}, \mathcal{F}} \langle \psi | \mathcal{E} \otimes \mathcal{F})(\phi) | \psi \rangle$$
$$= F_{\text{LO}}(\phi, \psi),$$

where the first equality is by definition and denoting the optimizers by $\mu, \{\mathcal{E}_\lambda\}, \{\mathcal{F}_\lambda\}$, the second is by linearity of

the Lebesgue integral, the inequality is because $\langle \psi | (\mathcal{E} \otimes \mathcal{F})(\phi) | \psi \rangle$ is a real number for any choice of local channels, the third equality is because μ is a probability measure that is now independent of the argument of the integral, and the final equality is by definition. This proves the reduction of LOSR to LO if the target state is pure.

Next, we bound the dimension of Σ. We want to consider $\max_{\mathcal{E}, \mathcal{F}} F(\psi, (\mathcal{E} \otimes \mathcal{F})(\phi))$. Without loss of generality, we assume the local spaces are "compressed" such that $d_{\text{in}} := \text{SR}(|\phi\rangle)$ so that $\mathcal{E}, \mathcal{F}$ both act on $\text{L}(\mathbb{C}^{d_{\text{in}}})$. We now show that without loss of generality we may restrict the output dimension of $\mathcal{E}, \mathcal{F}$ to be $d_{\text{out}} := \text{SR}(|\psi\rangle)$. This is just because we can project onto the support of the marginal of $|\psi\rangle$ on both local spaces, so we can restrict the local maps to this space. Formally, this can be seen as follows. Consider arbitrary $\mathcal{E}, \mathcal{F}$ and consider the target state (up to LU) $|\psi\rangle = \sum_i \sqrt{r_i}|i\rangle|i\rangle$. Define $\Pi_P := \sum_{i:r_i > 0}$, i.e., the projector onto the support of $\text{Tr}_B(\psi) = \text{Tr}_A(\psi)$, where the equality is up to the change in space. Note $\text{rank}(\Pi_P) = \text{Schmidt}(\psi)$. By construction, $(\Pi_P \otimes \Pi_P)|\psi\rangle = |\psi\rangle$. Therefore,

$$F(\psi, (\mathcal{E} \otimes \mathcal{F})(\phi)) = \langle \psi | (\mathcal{E} \otimes \mathcal{F})(\phi) | \psi \rangle$$
$$= \text{Tr}[|\psi\rangle\langle\psi| (\mathcal{E} \otimes \mathcal{F})(\phi)]$$
$$= \text{Tr}[\psi \Pi_P^{\otimes 2} (\mathcal{E} \otimes \mathcal{F})(\phi) \Pi_P^{\otimes 2}],$$

where in the first equality we have used item 4 of Proposition 1 and the other two use cyclicity of trace along with invariance of ψ under the projector. Now we can expand

$$\Pi_P^{\otimes 2}(\mathcal{E} \otimes \mathcal{F})(\phi)\Pi_P^{\otimes 2}$$
$$= \sum_{k,l} \Pi_P A_k \otimes \Pi_P B_k \phi A_k^\dagger \Pi_P \otimes B_l^\dagger \Pi_P$$
$$\equiv (\mathcal{E}_\Pi \otimes \mathcal{F}_\Pi)(\psi),$$

where $\{A_k\}, \{B_l\}$ are the Kraus operators of $\mathcal{E}, \mathcal{F}$, respectively, and $\mathcal{E}_\Pi, \mathcal{F}_\Pi$ are completely positive, trace non-increasing (CPTNI) maps defined by $\{\Pi_P A_k\}, \{\Pi_P B_l\}$, respectively. Note this equivalence holds as $(\Pi A_k)^\dagger = A_k^\dagger \Pi_P$ since $\Pi_P^\dagger = \Pi_P$ so it is CP and it is TNI because

$$\sum_k (\Pi_P A_k)^\dagger (\Pi_P A_k) = \sum_k A_k^\dagger \Pi_P A_k$$
$$\leqslant \sum_k A_k^\dagger \mathbb{1} A_k = \mathbb{1},$$

where we used $\Pi_P^2 = \Pi_P$ in the first equality, $\Pi_P \leqslant \mathbb{1}$ and that $\mathcal{E}$ is CP in the inequality, and that $\mathcal{E}$ is TP in the last inequality. An identical argument holds for $\mathcal{F}_P$. This proves the optimizer is achieved with CPTNI maps $\text{T}(\text{L}(\mathbb{C}^{d_{\text{in}}}), \text{L}(\mathbb{C}^{d_{\text{out}}}))$. Finally, we can lift $\mathcal{E}_P, \mathcal{F}_P$ to being CPTP, denoted $\widehat{\mathcal{E}}, \widehat{\mathcal{F}} \in \text{T}(\text{L}(\mathbb{C}^{d_{\text{in}}}), \text{L}(\mathbb{C}^{d_{\text{out}}}))$ by adding one Kraus operator, e.g., for $\mathcal{E}_P$ add the Kraus operator $Z \in \text{L}(\mathbb{C}^{d_{\text{in}}}, \mathbb{C}^{d_{\text{out}}})$ where $Z^\dagger Z = (\mathbb{1} - \sum_k A_k^\dagger \Pi A_k) \geqslant 0$ which always exists by definition of the space of positive-semidefinite operators. By linearity,

$$F(\psi, (\mathcal{E} \otimes \mathcal{F})(\phi)) = \text{Tr}[\psi(\mathcal{E}_\Pi \otimes \mathcal{F}_\Pi)(\phi)]$$
$$\leqslant \text{Tr}[\psi(\widehat{\mathcal{E}} \otimes \widehat{\mathcal{F}})(\phi)].$$

Therefore, without loss of generality, the optimal channels are $\mathcal{E}, \mathcal{F} \in \mathrm{C}(\mathbb{C}^{d_{\mathrm{in}}}, \mathbb{C}^{d_{\mathrm{out}}})$. Note this means that $\mathrm{Rank}(J_{\mathcal{E}}) \leqslant d_{\mathrm{in}} d_{\mathrm{out}}$ and likewise for $J_{\mathcal{F}}$.

We now derive the equation using the isometric representation of the channel [27]:

$$\max_{\mathcal{E}, \mathcal{F}} F(\psi, (\mathcal{E} \otimes \mathcal{F})(\phi))$$

$$= \langle \psi, (\mathcal{E} \otimes \mathcal{F})(\phi) \rangle$$

$$= \max_{V_1, V_2, |\zeta\rangle} |\langle \psi | \langle \zeta | (V_1 \otimes V_2) |\phi\rangle|^2$$

$$= \max_{U_1, U_2, |\zeta\rangle} \left| \langle \psi | \langle \zeta | (U_1 \otimes U_2) |\phi\rangle |0\rangle_{E_1} |0\rangle_{E_2} \right|^2$$

$$= \max_{U_1', U_2', |\zeta_{p'}\rangle} \left| \langle \psi | \langle \zeta_{p'} | (U_1' \otimes U_2') |\phi\rangle |0\rangle_{E_1} |0\rangle_{E_2} \right|$$

$$= \max_{P'} F((R \otimes P')^{\downarrow}, T_{\mathrm{embed}}^{\downarrow}),$$

where the second equality is because there exists an isometric representation of each channel which means $(V_1 \otimes V_2)|\phi\rangle$ is a pure quantum state, so we can apply Uhlmman's theorem to find a purification of $|\psi\rangle$ that saturates the bound, but as $|\psi\rangle$ is already pure, any purification will be a product state with a unit vector. The third line is because we can always convert an isometry into a unitary on the appropriately large space. The fourth line means that $\zeta_{p'} = \sum_{i'} \sqrt{p'(i)}|i\rangle|i\rangle$, which can always be achieved by local unitaries on the E_1 and E_2 spaces, which result in new unitaries on the other side but the same maximum. The final equality is just using Theorem 1 and we write T_{embed} to stress it is defined over the whole alphabet. Lastly, as we established bounds on the ranks of the local maps Choi matrices, we have bounds $E_1, E_2 \leqslant d_{\mathrm{in}} d_{\mathrm{out}}$, which justifies the maximum and tells us how large of a system we have to consider in the statement of the theorem. ∎

It is useful to see how this result works. It in effect shows the following equivalence of conversion distance when measured under fidelity:

$$\mathrm{F}\{|\phi\rangle \xrightarrow{\mathrm{LO}} |\psi\rangle]\} = \max_{|\zeta\rangle} \mathrm{F}\{|\phi\rangle \xrightarrow{\mathrm{LU}} |\psi\rangle \otimes |\zeta\rangle\}, \quad (10)$$

where here $\mathrm{F}\{a \rightarrow_{\mathcal{O}} b\}$ denotes the optimal conversion between a and b using maps included in set $\mathcal{O}$ according to fidelity. (See [20] for a discussion of conversion distance in terms of trace norm for general resource theories.) This statement can be viewed both by proof and via intuition as a special case of the isometric representation of a channel. Moreover, it is easy to see in this form how it handles our motivating example. Indeed, if the target state is $|\psi\rangle$ and the seed state is $|\psi\rangle \otimes |\zeta\rangle$, then clearly the maximizer is chosen by the ancillary state being $|\zeta\rangle$ and the local unitaries being trivial.

Example 3 (Tightened error bounds for zero communication). We now recall Example 1, which showed Proposition 2 could not show that any seed state $|\phi\rangle$ would require communication to be mapped to $|\psi\rangle = \sqrt{0.54}|00\rangle + \sqrt{0.02}|11\rangle + \sqrt{0.44}|22\rangle$. Here we calculate the error if the seed is a qutrit maximally entangled state, which we note is an example of the task of dilution. By Theorem 2,

$$F_{\mathrm{LOSR}}(|\psi\rangle, |\Phi_3^+\rangle) = \max_{P' \in \mathcal{P}([9])} F\left((P \otimes P')^{\downarrow}, \tfrac{1}{3}\mathbb{1}_{\mathbb{C}^3}\right)$$

$$= \tfrac{1}{3} \max_{P' \in \mathcal{P}^{\downarrow}([9])} F((P \otimes P')^{\downarrow}, \mathbb{1}_{\mathbb{C}^3}),$$

where we have used a slight abuse of notation as $\mathbb{1}_{\mathbb{C}^3}$ is embedded in a larger space. We note P' can be ordered as it will not change $(P \otimes P')^{\downarrow}$. By the definition of fidelity (1),

$$F((P \otimes P')^{\downarrow}, \mathbb{1}_{\mathbb{C}^3})$$

$$= \left(\max_{p' \in \mathcal{P}^{\downarrow}([9])} \sqrt{0.54 p'(1)} + \sqrt{0.44 p'(1)} + \sqrt{0.54 p'(2)} \right)^2,$$

where we have used $p'(1) \geqslant p'(2)$ and $0.54 > 0.44$, so these must be the largest three elements without loss of generality. As we are considering a maximization, it can only be increased by letting $p'(2) = 1 - p'(1)$. This allows us to parametrize in terms of $p' \in [0, 1]$. Taking the derivative with respect to p' and setting equal to zero, we find $p'^{\star} \approx 0.6014$. Plugging this back in, we get

$$F_{\mathrm{LOSR}}(|\psi\rangle, |\Phi_3^+\rangle) \approx \tfrac{1}{3}(1.59018)^2 \approx 0.8429 .$$

Thus, not only would communication be necessary, but without any, the fidelity can be far from one.

D. Relation between LO and LU strategies

The natural question given the previous theorems is if we can better understand the relationship between LO and LU strategies. We first show that LU and LO strategies are equivalent when either the target or the seed state is a two-qubit state.

1. LU and LO equivalence for two-qubit seed or target state

Proposition 8. Consider (possibly unnormalized) entangled two-qubit seed state $|\phi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$. Let the (possibly unnormalized) target entangled state be $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^{d'}$. Then the optimal noncommunicative strategy is the local unitary strategy.

Proof. First, we point out that it suffices to consider normalized distributions. This is because if we have unnormalized vectors, then, using the definition of fidelity,

$$F((R \otimes P')^{\downarrow}, T^{\downarrow}) = \mathrm{Tr}[R]\mathrm{Tr}[T]F((P \otimes P')^{\downarrow}, Q^{\downarrow}) ,$$

where $P := \mathrm{Tr}[R]^{-1}R$, $Q := \mathrm{Tr}[T]^{-1}T$. It follows since these scaling factors will appear for both the LU and LO cases (and thus cancel when comparing the values), the normalized case is sufficient.

Without loss of generality, $q^{\downarrow} = (q, 1 - q)$ where $q \geqslant \frac{1}{2}$ and $p^{\downarrow} = (p(1), p(2), \dots)$. Then the optimal local unitary strategy is $\sqrt{qp(1)} + \sqrt{(1 - q)p(2)}$. For any P' we can write $(p')^{\downarrow} = (p'(1), p'(2), \dots)$. The optimal CPTP strategy (up to a square) is of the form

$$\sqrt{qp(1)p'(1)} + \sqrt{(1 - q)\max\{p(1)p'(2), p(2)p'(1)\}} .$$

These values can only increase by assuming p' has two outcomes, so let us assume so without loss of generality and parametrize the distribution by $p' \in [\frac{1}{2}, 1]$ to obtain

$$\sqrt{qp(1)p'} + \sqrt{(1 - q)\max\{p(1)(1 - p'), p(2)p'\}} .$$

Moreover, note $p(2)p' < p(2)$ unless $p' = 1$, which is equivalent to the LU strategy, so the second entry in the maximization would be lower than the LU setting. Therefore, we focus on the remaining case. We are specifically interested

in when the following strict inequality holds:

$$\sqrt{qp(1)p'} + \sqrt{(1-q)p(1)(1-p')}$$
$$> \sqrt{qp(1)} + \sqrt{(1-q)p(2)}$$
$$\Leftrightarrow g(p') := \sqrt{qp(1)}(\sqrt{p'} - 1)$$
$$+ \sqrt{1-q}(\sqrt{p(1)(1-p')} - \sqrt{p(2)}) > 0.$$

Then $\frac{d}{dp'}g(p') = \frac{\sqrt{qp(1)}}{2\sqrt{p'}} + \frac{\sqrt{p(1)(1-q)}}{2\sqrt{1-p'}}$. It follows

$$\frac{\sqrt{qp(1)}\sqrt{1-p'}}{2\sqrt{p'}\sqrt{1-p'}} + \frac{\sqrt{p'}\sqrt{p(1)(1-q)}}{2\sqrt{1-p'}\sqrt{p'}} \geqslant 0$$
$$\Leftrightarrow \sqrt{qp(1)}\sqrt{1-p'} + \sqrt{p'}\sqrt{p(1)(1-q)} \geqslant 0$$
$$\Leftrightarrow \sqrt{q}\sqrt{1-p'} + \sqrt{p'}\sqrt{(1-q)} \geqslant 0$$
$$\Leftrightarrow \sqrt{F(Q^\downarrow, P'^\downarrow)} \geqslant 0,$$

where the first line is multiplying to get identical denominators, the second line is multiplying by the denominator, the third is dividing out $p(1)$, and the final is by the definition of square-root fidelity. Note the final inequality will always hold strictly unless $q \in \{0, 1\}$, i.e., the state is a product state, by item 1 of Proposition 1. If $q \in \{0, 1\}$, then the state is a product state which would contradict that we assume the state is entangled. Therefore, in our setting, $g(p')$ only increases over its interval $p' \in [0, 1]$. Thus, the optimal choice of p' is $p' = 1$, but in this case the value is $\sqrt{qp(1)} \leqslant \sqrt{qp(1)} + \sqrt{(1-q)p(2)}$, i.e., the optimal choice is lower bounding the optimal local unitary strategy. It follows this is never optimal. This completes the proof. ∎

2. LU and LO inequivalence for states with Schmidt rank greater than 2

If there is equivalence for two-qubit seed or target states, it is natural to ask if this property persists. One might expect that this is a special property of qubit systems as are found throughout quantum information science results. Indeed, generally this property does not hold, which we will prove via example.

Theorem 3. For seed and target state with Schmidt rank $\geqslant$ 3, the optimal LO strategy may be better than the optimal LU strategy.

Proof. We construct an example for Schmidt rank 3. By continuity of the fidelity, one can embed the target and seed in bigger spaces with arbitrarily small perturbations for it to hold in higher dimensions, which is why this is sufficient. Consider target state $|\psi\rangle = 0.85|00\rangle + 0.08|11\rangle + 0.07|22\rangle$ and seed state $|\phi\rangle = 0.45(|00\rangle + |11\rangle) + 0.1|22\rangle$. Then, the optimal LU strategy fidelity is

$$F(P^\downarrow, Q^\downarrow) = \left(\sqrt{0.45}(\sqrt{0.85} + \sqrt{0.08}) + \sqrt{0.1(0.07)}\right)^2$$
$$< 0.796.$$

In contrast, if we consider $P' = [0.55, 0.28, 0.17]$, then

$$F((P \otimes P')^\downarrow, Q^\downarrow)$$
$$= (\sqrt{0.45}\sqrt{0.4675} + \sqrt{0.45}\sqrt{0.238} + \sqrt{0.1}\sqrt{0.1445})^2$$
$$> 0.82.$$

As we maximize over P', the optimal LO strategy achieves a value that is strictly above the LU strategy. This completes the proof. ∎

E. Inefficiency of optimal LOSR fidelity and computable upper bounds

In the above we have constructed an example where the local operations strategy outperforms the local unitary strategy (though we have not shown what the strategy itself is). A natural question would then be how easy it is to solve for the optimal fidelity value or even a bound. By Theorem 1, we can conclude the optimal local unitary strategy is polynomial time to solve as all one needs to do is sort the Schmidt coefficients and calculate the fidelity. Indeed, one could solve for the ordering of the Schmidt coefficients using the linear program for sorting a vector.

In contrast, for optimizing LO strategies, we have no such luck. In effect this is because there are two things to optimize over at once. Indeed, recall from Theorem 2 that

$$F_{\text{LO}}(|\psi\rangle, |\phi\rangle) = \max_{P' \in \mathcal{P}(\Sigma)} F((P \otimes P')^\downarrow, Q^\downarrow).$$

Then the problem is that one must first tensor P onto variable P' and then reorder the vector. One cannot even in general order an optimization variable, which we will refer to as "sorting," as sorting is in general nonconvex. In sorting a vector using a linear program, one relaxes to bistochastic channels and considers a linear function so that the optimizer is an extreme point which by the Birkhof–von Neumann theorem is a specific permutation. However, we are many levels of involvement above that: we want the distribution P' such that its product distribution $P \otimes P'$ when sorted optimizes the fidelity with $Q^\downarrow$. Therefore, we need to optimize over P' and the permutation at the same time. It is not clear that we can actually relax to bistochastic strategies because of the joint concavity of fidelity. That is to say, for any bistochastic channel $\mathcal{E}$,

$$F(\mathcal{E}(P \otimes P'), Q^\downarrow) = F\left(\sum_\pi r(\pi)V_\pi(P \otimes P'), Q^\downarrow\right)$$
$$\geqslant \sum_\pi F(r(\pi)V_\pi(P \otimes P'), r(\pi)Q^\downarrow)$$
$$= \sum_\pi r(\pi)F(V_\pi(P \otimes P'), Q^\downarrow),$$

where the first line is Birkhoff–von Neumann theorem, the second is joint concavity using $Q^\downarrow = \sum_\pi r(\pi)Q^\downarrow$ as r is a probability distribution, and the last line is because $F(\lambda P, Q) = \lambda F(P, Q) = F(P, \lambda Q)$. Thus, any bistochastic channel *may* strictly do better than the average of its extreme points. Moreover, even if we could optimize over bistochastic channels, we would have a nonconvex objective function as the bistochastic channel, an optimization variable, would be applied to $P \otimes P'$ which is also partially an optimization variable.

Given the above, it seems likely the best option if one were to try and find a (near) optimum would be to use gradient descent from random initial P', realizing it will only work locally and will break down at "kinks" where the ordering

changes. Otherwise, more sophisticated nonconvex optimization techniques might be used.

Computable upper bound methods. Perhaps even worse than our inability to calculate the exact fidelity, is that it is not clear in general how to determine good bounds. Certainly, we have the following result.

Theorem 4. Unless the target state is (up to local unitaries) $|\psi\rangle = |\phi\rangle \otimes |\zeta\rangle$ for some pure state $|\zeta\rangle$ where $|\phi\rangle$ is the seed state, there exists $\varepsilon > 0$ such that there does not exist local operations that will take $|\phi\rangle$ to $|\psi\rangle$.

Proof. Theorem 2 states that

$$F_{\text{LOSR}}(|\psi\rangle, |\phi\rangle) = \max_{P' \in \mathcal{P}(\Sigma)} F((R \otimes P')^{\downarrow}, T_{\text{embed}}^{\downarrow}) .$$

Item 1 of Proposition 1 states that fidelity between two normalized states (and thus distributions) is one if and only if the two arguments are the same. Thus, the maximization obtains one if and only if there exists $P' \in \mathcal{P}(\Sigma)$ such that $(R \otimes P')^{\downarrow} = T_{\text{embed}}^{\downarrow}$. This means the Schmidt coefficients of $|\phi\rangle$ are $\{r_i p'_k\}_{(i,k)}$. That is (up to local unitaries) $|\phi\rangle = \sum_{(i,k)} r_i p'_k |(i, k)\rangle |(i, k)\rangle$. Defining local isometry $U_{A \to A_0 A_1} |(i, k)\rangle_A = |i\rangle_{A_0} |k\rangle_{A_1}$, we have

$$(U \otimes U)|\psi\rangle = \sum_{i,k} r_i p'_k (|i\rangle_{A_0} |k\rangle_{B_0} |i\rangle_{A_1} |k\rangle_{B_1})$$

$$= \left(\sum_i r_i |i\rangle_{A_0} |i\rangle_{B_0} \right) \otimes \left(\sum_k p_k |k\rangle_{A_1} |k\rangle_{B_1} \right)$$

$$=: |\psi\rangle \otimes |\zeta\rangle,$$

where we defined $|\zeta\rangle$ as the second state. This completes the proof. ∎

The above theorem, while derived from a very different strategy than Proposition 2, does not seem to give us much more information as to at what point communication is necessary. What we would want to efficiently improve this would be to establish upper bounds on the equation given in Theorem 2 that have a closed form that does not depend on P'. One option is to use the data processing inequality for fidelity. This can be seen in the following proposition.

Proposition 9. Consider (possibly unnormalized) target state $|\psi\rangle$ and seed state $|\phi\rangle$ with corresponding Schmidt distributions p, q, respectively. If $p_{\max} \leqslant q_{\max}$, then

$$F_{\text{LO}}(|\psi\rangle, |\phi\rangle) \leqslant F(\overline{p}, \overline{q}),$$

where $\overline{p} = p_{\max} |0\rangle\langle 0| + (1 - p_{\max})|1\rangle\langle 1|$ and likewise for $\overline{q}$.

Proof. Without loss of generality let d be the maximum local dimension. Let $\mathcal{E}(\cdot) = |0\rangle\langle 0||0\rangle\langle 0| + \sum_{i \in \{1, d-1\}} |1\rangle\langle i||i\rangle\langle 1|$. That is, $\mathcal{E}$ coarse grains a probability distribution to the Bernoulli distribution with its first element untouched and the sum of all the others as the other outcome. Then using data processing of fidelity (item 3 of Proposition 1),

$$\max_{P' \in \mathcal{P}(\Sigma)} F((P \otimes P')^{\downarrow}, Q^{\downarrow}) \leqslant \max_{P' \in \mathcal{P}(\Sigma)} F(\mathcal{E}[(P \otimes P')^{\downarrow}], \mathcal{E}(Q^{\downarrow}))$$

$$= \max_{p' \in [0,1]} F(\widetilde{P}(p'), \mathcal{E}(Q^{\downarrow})),$$

where $\widetilde{P}(p') := p_{\max} p' |0\rangle\langle 0| + (1 - p_{\max} p')|1\rangle\langle 1|$ and $\mathcal{E}(Q^{\downarrow}) = q_{\max} |0\rangle\langle 0| - (1 - q_{\max})|1\rangle\langle 1|$. Now note that by

assumption $p_{\max} \leqslant q_{\max}$. As the fidelity will only decrease as $p_{\max} p'$ moves away from $q_{\max}$, the optimal choice is $p' = 1$. This completes the proof. ∎

The problem with the above bound is that there will be cases where $p_{\max} > q_{\max}$. Why the inequality in the other direction was required was to know for a fact what element of p was relevant, namely, $p_{\max}$ and that any choice of $p' \neq 1$ would be suboptimal. In general this strategy would require $q^{\downarrow}(j)$ is sufficiently large relative to $p^{\downarrow}(j)$. This can be determined in some cases. Here we provide a simple example.

Example 4. Let

$$p^{\downarrow} = [3/4, 1/8, 1/8]^{\text{T}}, \quad q^{\downarrow} = [1/2, 1/2]^{\text{T}}.$$

Then $(p \otimes p')^{\downarrow}[1 : 2] = p'(1)[\frac{3}{4}, \frac{1}{8}]^{\text{T}}$, and so we can coarse grain on the second element to obtain $\overline{P}(p') = 1/8p'|0\rangle\langle 0| + (1 - 1/8p')|1\rangle\langle 1|$ and $\overline{Q} = Q^{\downarrow}$. Then as $1/8p' < \frac{1}{2}$, the upper bound is $F(\frac{1}{8}|0\rangle\langle 0| + \frac{7}{8}|1\rangle\langle 1|, \frac{1}{2}\mathbb{1}) \approx 0.83$.

The above shows that while data processing can be sufficient in certain cases, it does not provide an easy general method. Another common alternative in quantum information theory is semidefinite relaxations of optimization problems because semidefinite programs are efficient to evaluate. In Appendix B, we establish the following upper bound and show it may be expressed as a semidefinite program, which, as everything is in terms of probability distributions, is due to the nonlinearity of fidelity and nothing particularly quantum.

Theorem 5. Consider (possibly unnormalized) target state $|\psi\rangle$ and (possibly unnormalized) seed state $|\phi\rangle$. Let $\text{SR}(\psi) = d$ and $\text{SR}(\phi) = d'$. Define $A = \mathbb{C}^d$, $B = \mathbb{C}^{dd'}$. Then,

$$F_{\text{LOSR}}(|\psi\rangle, |\phi\rangle) \leqslant \max \; F(R, Q_{\text{embed}}^{\downarrow}) \text{s.t. } \text{Tr}_B[R] = P^{\downarrow}$$

$$\times R \in \mathcal{P}^{\downarrow}(d^2 d'), \tag{11}$$

where P and Q are the distributions defined by $|\psi\rangle$ and $|\phi\rangle$'s Schmidt coefficients, respectively. Moreover, this admits the following simple semidefinite program over the reals:

$$\max_{i \in [d^2 d']} \sum x(i) \text{s.t. } \begin{pmatrix} \text{diag}(r) & \text{diag}(x) \\ \text{diag}(x) & \text{diag}(q_{\text{embed}}^{\downarrow}) \end{pmatrix} \succeq 0$$

$$\times \text{Tr}_B[\text{diag}(r)] = P^{\downarrow} r \in \mathcal{P}^{\downarrow}([d^2 d]) x \in \mathbb{R}^{d^2 d'}. \tag{12}$$

Physically, this relaxation may be seen as relaxing the isometric representation of the optimal LOSR strategy to one where one allows the ancillary environment start off entangled with the local system. Mathematically, this is not too loose because we require this entangled pure quantum state has a notion of "local Schmidt coefficients" that pertain to the original target state, although this physically does not seem to have a clean interpretation. Nonetheless, we can see that (11) will not achieve unity unless there exists a joint distribution $Q = R$, which would require $Q_{\text{embed}}^{\downarrow}$ to have $P^{\downarrow}$ as it is marginal, which seems highly restrictive. Therefore, (11) should provide an upper bound that is nontrivial.

V. MANY-COPY PURE STATE CONVERSION WITH ZERO COMMUNICATION

Having established what happens for single copies, we consider many copies. We provide two motivations for doing this. First, we note that it is not clear what the limiting behavior will be even in the LU setting. A reader may recall from other works that the fidelity is multiplicative so if $F(P, Q) < 1$, then $\lim_{n\to\infty} F(P^{\otimes n}, Q^{\otimes n}) = \lim_{n\to\infty} F(P, Q)^n \to 0$. However, we lose the multiplicativity as we are considering $\lim_{n\to\infty} F((P^{\otimes n})^{\downarrow}, (Q^{\otimes n})^{\downarrow})$. This issue is further aggravated if we consider local operations and the ancillary variable.

The second motivation is that what was initially considered in the literature, albeit with LOCC [33], was the conversion of many copies of states. A particular focus in the referenced work and subsequent ones is the case where either the target or seed state is the maximally entangled state, known as distillation and dilution, respectively. With LOCC, we know there are "rates" in the conversions. By [6] along with previous results in this work, we would not expect there to be non-negative rates without the communication assuming the error is required to be vanishing, i.e., $\varepsilon \to 0$.

In this section we establish convex optimization problems for dilution and distillation in the zero-communication setting. These results are established in terms of the not-actually-a-norm $\|\cdot\|_{(k,1/2)}$, which we remind the reader is the (k, p)-norms extended to $p < 1$ introduced in Sec. III with the choice of $p = \frac{1}{2}$. We also look at the limiting behavior as the number of copies grows. In particular, we find a closed form when trying to convert n-fold two-qubit states to a different n-fold two-qubit state. Moreover, we prove the fidelity goes to zero in this case. We discuss the extension of this to entangled states with larger Schmidt rank.

1. Dilution under local operations

We begin by determining the limits of dilution. For intuition, we begin with local unitaries where there is no optimization. Recall that the Schmidt coefficients of the maximally entangled state are all $\sqrt{d^{-1}}$, so they correspond to the maximally mixed distribution under our bijection between Schmidt coefficients and probability distributions.

Proposition 10. For local unitary strategies the optimal dilution fidelity is given by

$$F_{\text{LU}}\big(|\psi\rangle, |\Phi_d^+\rangle^{\otimes n}\big) = d^{-n}\|P\|_{(d^n,1/2)} .$$

Proof. Generally, if $|\psi\rangle \neq |\Phi_d^+\rangle$,

$$1 > F\big(P^{\downarrow}, \pi_d^{\otimes n \downarrow}\big) = F\big(P^{\downarrow}, \pi_d^{\otimes n}\big) = \left[d^{-n/2}\sum_{i\in[d^n]}\sqrt{P^{\downarrow}(i)}\right]^2$$

$$= d^{-n}\left[\sum_{i\in[d^n]}\sqrt{P^{\downarrow}(i)}\right]^2 = d^{-n}\|P\|_{(d^n,1/2)},$$

where the first equality is because $\pi_d^{\otimes n}$ is invariant under ordering, the second is using the definition of fidelity and that $\pi_d^{\otimes n}$ has uniform coefficients, and the final equality is the definition of the (k, p)-norms. In particular, note we have dropped the sorting. ∎

We remark we could have set $|\phi\rangle = |\phi'\rangle^{\otimes m}$ to get a trade-off, but this does not seem to provide any insight. Just as in

the one-shot setting, we know the above result is not as useful in general because it cannot throw out resources, so we now present the general result.

Proposition 11. The optimal fidelity of converting n d-local dimensional EPR pairs to $|\psi\rangle$ under local operations is given by

$$F_{\text{LO}}(|\psi\rangle, |\Phi_d^+\rangle^{\otimes n}) = d^{-n}\max_{P'\in\mathcal{P}(\Sigma)}\|(P\otimes P')\|_{(d^n,1/2)} ,$$

where $\|\cdot\|_{(k,p)}$ is (k, p)-norm generalized to $p \geqslant 0$. Moreover, for fixed n, this is a convex optimization problem.

Proof. Starting from the result of Theorem 2,

$$\max_{P'\in\mathcal{P}(\Sigma)} F((P\otimes P')^{\downarrow}, (\pi_d^{\otimes n})^{\downarrow})$$

$$= \max_{P'\in\mathcal{P}(\Sigma)} F((P\otimes P')^{\downarrow}, \pi_d^{\otimes n})$$

$$= \left[\frac{1}{d^{n/2}}\max_{P'\in\mathcal{P}(\Sigma)}\sum_{i\in[d^n]}\sqrt{(P\otimes P')^{\downarrow}(i)}\right]^2 \quad (\star)$$

$$= \frac{1}{d^n}\max_{P'\in\mathcal{P}(\Sigma)}\|P\otimes P'\|_{(d^n,1/2)},$$

the first inequality is invariance of $\pi_d^{\otimes n}$ under sorting, the second is definition of fidelity and that each element of $\pi_d^{\otimes n}$ is the same, the last is the definition of (k, p)-norm extended to $p \geqslant 0$.

To show this is a convex optimization problem, note that $\Phi_P(\cdot) := P \otimes \cdot$ is linear, $-\sqrt{\cdot}$ is operator convex, and the sum of the k largest eigenvalues of a PSD P, which we will denote $\Sigma_k(P)$ is convex. Thus, starting from $(\star)$,

$$\left[d^{-n/2}\max_{P'\in\mathcal{P}(\Sigma)}\sum_{i\in[d^n]}\sqrt{P\otimes P'}^{\downarrow}(i)\right]^2$$

$$= \left[-d^{-n/2}\min_{P'\in\mathcal{P}(\Sigma)}\Sigma_{d^n}(-\sqrt{\Phi_P(P')})\right]^2,$$

where we have used $\max_{x\in\mathcal{C}} f(x) = -\min_{x\in\mathcal{C}} -f(x)$ and our definitions. Then ignoring the $-d^{-n/2}$ factor and the square, the optimization problem is over the probability simplex, which is a convex subset of the positive-semidefinite matrices, and the objective function is convex over the positive-semidefinite cone as $-\sqrt{\Phi_P(\cdot)}$ is operator convex and Σ_{d^n} is a convex function over the space of Hermitian operators. This completes the proof. ∎

Unfortunately, while this gives computable bounds, it is not clear how one could determine the optimal value analytically.

2. Distillation under local operations

We now present the same results in the distillation case, where we take some state to many EPR states. For completeness, we state the local unitaries case.

Proposition 12. The fidelity of distillation under local unitaries and zero communication is given by

$$F_{\text{LU}}(|\Phi_d^+\rangle^{\otimes m}, |\psi\rangle^{\otimes n}) = d^{-m}\|P^{\otimes n}\|_{|S|,1/2} ,$$

where $S = [\min\{d^m, \text{rank}(P)^n\}]$.

Proof. The proof is effectively identical to the dilution case by symmetry of the fidelity. ∎

In contrast to the local unitary case, the symmetry is broken when one considers local operations.

Theorem 6. For fixed d, m, n the optimal fidelity for distillation under local operations is given by

$$F_{\text{LO}}(|\Phi_d^+\rangle^{\otimes m}, |\psi\rangle^{\otimes n}) = d^{-m}\left[\min_{P' \in \mathcal{P}^\downarrow(\Sigma)} -\sum_{i \in \mathcal{I}} \alpha_i \sqrt{p'(i)}\right]^2,$$

where $\mathcal{P}^\downarrow(\Sigma)$ is the set of decreasing distributions as defined in Sec. III, $\mathcal{I} \equiv [\lceil \text{rank}(P)^n/d^m \rceil]$, and $\alpha_i := \sum_{j \in [(i-1)d^m:\min\{i\cdot d^m, \text{rank}(P)^n\}]} \sqrt{p_n^\downarrow(i)}$. Note the minimization is a convex optimization program.

Proof. Yet again, we use the square-root fidelity and then take the square at the end. Then, using Theorem 2, we have

$$F_{\text{LO}}((\Phi_d^+)^{\otimes m}, \psi^{\otimes n})$$
$$= \max_{P' \in \mathcal{P}(\Sigma)} F((\pi_d^{\otimes m} \otimes P')^\downarrow, (P^{\otimes n})^\downarrow)$$
$$= \left[\max_{P' \in \mathcal{P}(\Sigma)} \sum_{i \in S} \sqrt{(\pi_d^{\otimes m} \otimes p')^\downarrow(i)}\sqrt{p_n^\downarrow(i)}\right]^2.$$

Next, note

$$\left(\pi_d^{\otimes m} \otimes P'\right)^\downarrow = d^{-m/2}\sum_{i' \in \Sigma} p^\downarrow(i')\mathbb{1}_{\mathbb{C}^{d^m}},$$

where we have just used that $\pi_d^{\otimes m}$ is invariant under ordering. It follows that if we let $\mathcal{I} \equiv [\lceil \text{rank}(P)^n/d^m \rceil]$, we can rewrite

$$F_{\text{LO}}((\Phi_d^+)^{\otimes m}, \psi^{\otimes n}) = d^{-m}\left[\max_{P' \in \mathcal{P}(\Sigma)} \sum_{i \in \mathcal{I}} \sqrt{(p')^\downarrow(i)}\right.$$
$$\left. \times \sum_{j \in [(i-1)d^m:\min\{id^m, \text{rank}(P)^n\}]} \sqrt{p_n^\downarrow(i)}\right]^2.$$

Now first define $\alpha_i := \sum_{j \in [(i-1)d^m:\min\{id^m, \text{rank}(P)^n\}]} \sqrt{p_n^\downarrow(i)}$ as these coefficients may be precomputed. Second, note that the probability simplex restricted to descending distributions $\mathcal{P}^\downarrow(\Sigma)$ is itself convex as $r_\lambda^\downarrow := \lambda p^\downarrow + (1-\lambda)q^\downarrow$ satisfies

$$\lambda p^\downarrow(i) + (1-\lambda)q^\downarrow(i) \geq \lambda p^\downarrow(i+1) + (1-\lambda)q^\downarrow(i),$$

for all $i \in [|r|]$. Thus we have

$$F_{\text{LO}}(|\Phi_d^+\rangle^{\otimes m}, |\psi\rangle^{\otimes n}) = \left[-d^{-m}\min_{P' \in \mathcal{P}^\downarrow(\Sigma)} -\sum_{i \in \mathcal{I}} \alpha_i \sqrt{p'(i)}\right]^2.$$

The minimization is a convex optimization problem because if we consider $f(p') := -\sum_i \alpha_i \sqrt{p'(i)}$, then its Hessian is $\nabla^2 f = \sum_i [\alpha_i/4 p'(i)^{-3/2}]|i\rangle\langle i|$, which is positive semidefinite on the interior of the probability simplex [i.e., when $p'(i) > 0$ for all i]. This completes the proof. ∎

3. Two-qubit setting

We have now seen that even in the basic dilution and distillation setting, while we can determine convex optimization programs, we cannot seem to get clean analytic results. In this section we consider an even more tractable setting to attempt to resolve this: many-copy two-qubit seed and target states. We show in this setting under certain assumptions the local

unitary strategy is optimal and lobby this to show in particular that the optimal fidelity of converting n copies of $|\phi\rangle$ to n copies $|\psi\rangle$ goes to zero as n goes to infinity. We note that this setting is more manageable because we effectively only have to reason about Bernoulli distributions.

Lemma 3. Given Bernoulli distribution $P = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$, then $P^{\otimes n}$ is such that the sequence x^n with $(n-k)$ zeros has probability $p^{n-k}(1-p)^k$. Moreover, there are $\binom{n}{k}$ sequences with probability $p^k(1-p)^{n-k}$ and the same for $p^{n-k}(1-p)^k$.

Proof. The claim that x^n with $(n-k)$ zeros has probability $p^{n-k}(1-p)^k$ is straightforward. The second point actually just follows from the fact there are $\binom{n}{k}$ sequences with k zeros, which could be proven by induction in a straightforward manner. ∎

We can now use the above lemma along with Theorem 1 to get the optimal LU fidelity as a function of the number of copies n.

Corollary 2. Consider entangled states $|\psi\rangle, |\phi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$. Then,

$$F_{\text{LU}}(\psi^{\otimes n}, \phi^{\otimes n}) = \sum_{k \in [n]} \binom{n}{k}(pq)^{(n-k)/2}[(1-p)(1-q)]^{k/2}.$$

Proof. By Theorem 1 we can reduce to the Bernoulli distributions from the Schmidt coefficients $|\psi\rangle^{\otimes n} \mapsto P^{\otimes n}$, $|\phi\rangle^{\otimes n} \mapsto Q^{\otimes n}$. Since these are Bernoulli distributions, if we assume without loss of generality $p \geq (1-p)$, we can order the probabilities simply by the exponent, e.g., $p^{j-k}(1-p)^k \geq p^{j-k-k'}(1-p)^{k+k'}$ for any $0 \leq k' \leq j-k$. Moreover, the cardinality of each set of sequences will be the same for both $P^{\otimes n}$ and $Q^{\otimes n}$ because $|\psi\rangle, |\phi\rangle$ are only entangled if their Schmidt rank is 2. Therefore,

$$F((P^{\otimes n})^\downarrow, (Q^{\otimes n})^\downarrow) = \sum_{k \in [n]} \binom{n}{k}(pq)^{(n-k)/2}[(1-p)(1-q)]^{k/2},$$
(13)

where the sum is over the number of zeros in the string, the cardinality was proven in the previous lemma, and the last term is just a rewriting of $\sqrt{p^{n-k}(1-p)^k}\sqrt{q^{n-k}(1-q)^k}$. ∎

We note it is straightforward to generalize the above result to the case where you have the number of states differ between the seed and the target, but the form would be ugly as one would need to count how many sequences of a given probability there are and keep track of this in the sum. Indeed, at this point the problem is elaborate enough that there is no advantage with dealing with two-qubit states as it is a question of the type classes [34]. We state this as a remark.

Remark 1. Consider states $|\psi\rangle, |\phi\rangle$, respectively, with ordered probability distributions corresponding to their Schmidt coefficients P and Q, respectively. $F_{\text{LU}}(|\psi\rangle^{\otimes n}, |\phi\rangle^{\otimes m})$ can be computed. This is because the probability of a given sequence drawn in i.i.d. form from a distribution has a closed form [34, Theorem 11.1.2]. It follows that as long as one determines the type of classes exactly and takes into account that the sizes of the type of classes may differ between P and Q, the computation is possible, albeit tedious.

Rather than dealing with the computational nightmare of generalizing beyond two-qubit states, we now show that the term in Corollary 2 always goes to zero as n goes to infinity.

Proposition 13. Consider entangled states $|\psi\rangle, |\phi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$:

$$\lim_{n\to\infty} F_{\text{LU}}(|\phi\rangle^{\otimes n}, |\psi\rangle^{\otimes n}) = 0 \,.$$

Proof. Let the probability distributions corresponding to their Schmidt coefficients be parametrized by $p \geqslant \frac{1}{2}$ and $\frac{1}{2} \leqslant q = p + \varepsilon$ where $\varepsilon \in [-\frac{1}{2}, \frac{1}{2}]$. Therefore, using Corollary 2, we have

$$F_{\text{LU}}(|\psi\rangle^{\otimes n}, |\phi\rangle^{\otimes n}) = \sum_{k\in[n]} \binom{n}{k} (p^2 + p\varepsilon)^{(n-k)/2}$$
$$\times [(1-p)^2 - \varepsilon(1-p)]^{k/2}. \quad (14)$$

Now note $p^2 + p\varepsilon = pq < 1$ as otherwise both states would be product. Define $\alpha := (p^2 + p\varepsilon)^{1/2} < 1$. Then we have

$$\binom{n}{k} (p^2 + p\varepsilon)^{(n-k)/2} [(1-p)^2 - \varepsilon(1-p)]^{k/2}$$
$$\leqslant \left(\frac{ne}{k}\right)^k \alpha^{n-k} [(1-p)^2 - \varepsilon(1-p)]^{k/2}$$
$$= \left(\frac{e}{k}\alpha^{-1}\right)^k [(1-p)^2 - \varepsilon(1-p)]^{k/2} n^k \alpha^n$$
$$= O[\text{poly}(n)]O[\exp(-n)]$$
$$\to 0,$$

where the inequality uses a standard upper bound on the binomial coefficient, in the first equality we have grouped terms by scaling. The second equality uses asymptotic notation, where we remind the reader $g(n) = O[f(n)]$ means for sufficiently large n, $g(n) \leqslant Cf(n)$ for some constant C and poly(n) [respectively exp(n)] denote the sets of functions polynomial (respectively exponential) in integer n. To make this conversion, we have used that everything but α^n is a polynomial in n and that $\alpha < 1$, so α^n scales inverse exponentially in n. The limiting factor is then because an inverse exponential times a polynomial goes to zero. We also remark that the term where $k = n$ will also go to zero as $[(1-p)^2 - \varepsilon(1-p)]^{k/2}$ will go to zero as k goes to infinity as its magnitude will be bounded by 1.

Therefore, each term in the sum goes to zero as n goes to infinity, so the entire sum will go to zero. This completes the proof. ∎

We note our proof tells us nothing about the scaling as a function of the difference between p and q nor does it tell us how fast it goes to zero compared to $F(P^{\otimes n}, Q^{\otimes n})$. These are shown numerically for specific cases in Fig. 4.

It is then natural to ask if what we have seen so far is something special to local unitaries. We show that under sufficient conditions, just like in the single-copy case, when two-qubit seed states are involved, local unitary strategies are optimal.

Theorem 7. Let $|\psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ and the target state be $|\psi\rangle^{\otimes n}$. Let the seed state $|\phi\rangle$ satisfy $\text{SR}(|\phi\rangle) \leqslant n\text{SR}(|\psi\rangle)$. Then the optimal local operations strategy is the optimal local unitary strategy.
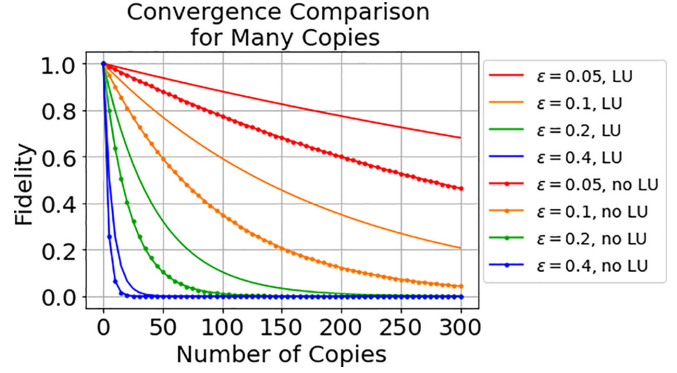

FIG. 4. Comparison of $F(|\psi\rangle^{\otimes n}, |\phi\rangle^{\otimes n})$ and $F_{\text{LU}}(|\psi\rangle^{\otimes n}, |\phi\rangle^{\otimes n})$ as a function of how different the Schmidt coefficients are and as a function of the number of copies. Here we consider many copies of $|\psi\rangle = \sqrt{0.55}|0\rangle|0\rangle + \sqrt{0.45}|1\rangle|1\rangle$ being converted into the same number of copies of $|\phi\rangle = \sqrt{q}|0\rangle|0\rangle + \sqrt{1-q}|1\rangle|1\rangle$ where $q := p + \varepsilon$ for choices of ε denoted in the legend. Dotted lines denote when there are no local unitaries applied to reorder the Schmidt coefficients where as straight lines are the case where the local unitary strategy is applied.

Proof. By Theorem 2,
$$F_{\text{LO}}(|\psi\rangle^{\otimes n}, |\phi\rangle) = \max_{P'\in\mathcal{P}(\Sigma)} F((P^{\otimes n}\otimes P')^{\downarrow}, Q^{\downarrow})$$
$$= \sum_{i\in|Q|} \sqrt{Q^{\downarrow}(i)}\sqrt{(P^{\otimes n}\otimes P')^{\downarrow}(i)}.$$

We will show that P' should be the delta distribution. If $p \neq \frac{1}{2}$, $p'(1) < 1$, then for any $0 \leqslant k \leqslant n$, we have the inequalities
$$p^{n-k}(1-p)^k > p^{n-k}(1-p)^k p'(1)$$
$$> p^{n-k}(1-p)^k p'(2)$$

and
$$p^{n-k}(1-p)^k > p^{n-k}(1-p)^k p'(1)$$
$$> p^{n-(k+1)}(1-p)^{k+1} p'(1)$$
$$> p^{n-(k+1)}(1-p)^{k+1} p'(2).$$

As square root is a monotone, this holds when we take the square root. Note that by assumption $P^{\otimes n}$ has enough entries by itself for there to be one corresponding to each $q^{\downarrow}$. Therefore, given the inequalities above, it follows if $p'(1) \neq 1$, each term in the sum only decreases. Therefore, $p'(1)$ is optimal for every n and k. Thus, when $p \neq \frac{1}{2}$, the optimal value is obtained by P' being a delta distribution, which means it is equivalent to the local unitary strategy.

Finally, if $p = \frac{1}{2}$, then $p^{n-k}(1-p)^k = 2^{-n}$ for all k. Therefore, if $p'(1) < 1$, the inequalities simplify for all $0 \leqslant k \leqslant n$:
$$p^{n-k}(1-p)^k p'(1) = p^{n-(k+1)}(1-p)^{k+1} p'(1)$$
$$> p^{n-(k+1)}(1-p)^{k+1} p'(2)$$

and
$$p^{n-k}(1-p)^k p'(1) > p^{n-k}(1-p)^k p'(2).$$

Again because each q term is paired up already, this means if $p'(1) \neq 1$, the value decreases. Therefore, we again conclude

the optimal strategy is the LU strategy. This completes the proof. ∎

We note that a trivial example of why we need the Schmidt rank constraint in the previous theorem is our original example for the advantage of LO strategies: if $|\phi\rangle^{\otimes n+\ell}$ where $\ell \geqslant 1$, then there is a better LO strategy than an LU strategy. Finally, we note it immediately follows from these previous results that

Corollary 3. If $|\phi\rangle, |\psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ are both entangled, then

$$\lim_{n\to\infty} F_{\text{LO}}(|\psi\rangle^{\otimes n}, |\phi\rangle^{\otimes n}) = 0 .$$

VI. ON GENERALIZED EMBEZZLING CONVERSION

We now have established a rather robust theory of pure state transformations under local operations. It is natural to return to the topic of conversion of one state to another using an ancillary entanglement, i.e., catalytic transformations and their embezzling relaxation. Of course, it is immediate from our results so far that we know the optimization program that determines the optimal pure embezzling state as we state in the following proposition.

Proposition 14. For any Schmidt rank d, the optimal pure state embezzler for state conversion $|\phi\rangle$ to $|\psi\rangle$ is the quantum state $|\zeta\rangle = \text{vec}(\sqrt{R})$ that is determined via the optimization

$$\max_{R \in \mathcal{P}(d), P' \in \mathcal{P}(\Sigma)} F((P \otimes P')^{\downarrow}, (Q \otimes R)^{\downarrow}). \quad (15)$$

Proof. This immediately follows from the input being $|\phi\rangle \otimes \text{vec}(\sqrt{R})$ and then applying Theorem 2. Note this means $|\Sigma|$ scales as function of d. ∎

However, as we have already addressed, even without a free variable for the embezzling state, the optimization in Theorem 2 seems unmanageable directly. While in principle one could use the relaxation in Theorem 5 to obtain efficient upper bounds, it is less obvious how often these will be nontrivial given that R is a free variable.

The next most natural setting would be that of approximate catalytic state conversion under local unitaries, i.e., we consider transformations of the form

$$|\phi\rangle|\zeta\rangle \approx_{\varepsilon}^{\text{LU}} |\psi\rangle|\zeta\rangle, \quad (16)$$

where $|\zeta\rangle$ is an embezzling state and $\approx_{\varepsilon}^{\text{LU}}$ denotes reversible equivalence under local unitary transformation up to error ε as measured under fidelity. This may be seen as a generalization of traditional embezzlement where $|\phi\rangle = |0\rangle_A|0\rangle_B$ and $|\zeta\rangle = |\mu(n)\rangle$.[2]

Now as noted in the background, embezzling is known to be in effect optimal for sufficiently small ε. It follows for sufficiently small error $\varepsilon > 0$, the strategy that embezzles out the seed state and then embezzles in the target state is roughly optimal, i.e.,

$$|\phi\rangle|\mu(n)\rangle \approx_{\varepsilon}^{\text{LU}} |0\rangle|0\rangle|\mu(n)\rangle \approx_{\varepsilon}^{\text{LU}} |\psi\rangle|\mu(n)\rangle \quad (17)$$

is effectively optimal where we remind the reader $|\mu(n)\rangle$ is the van Dam–Hayden embezzling family n pertains to the

Schmidt rank of the given state in the family (see Proposition 3). Nonetheless, we may explore at what point this becomes necessary.

Using Theorem 1, we know the optimal strategy is given by[3]

$$\max_{R \in \mathcal{P}(d)} F((P \otimes R)^{\downarrow}, (Q \otimes R)^{\downarrow}).$$

Even in the case $P, Q, R \in \mathcal{P}(2)$ this technically cannott be solved using gradient methods as one has to sort the $p(1-r)$ and $(1-p)r$ terms of $p \otimes r$ and likewise for $q \otimes r$. Nonetheless, it is hopefully clear that $r \in [\min\{p, q\}, \max\{p, q\}]$, as it is trying to make the distributions be more similar. Nonetheless, this issue will only grow in difficulty with the dimension and it is unclear how one would prove an ansatz is optimal in general. Therefore, we provide two-qubit examples which characterizes the general insights.

Example 5 (Resource gap between van Dam–Hayden embezzling state and optimal embezzler). Consider Bernoulli distributions P, Q, R parametrized by $p = 0.5, q = 0.7$ and we leave r unspecified for now. In other words, one of the states is the maximally entangled states and the other is, up to local unitaries, $\sqrt{0.7}|00\rangle + \sqrt{0.3}|11\rangle$. Therefore, depending on which way one runs the transformation, we are considering entanglement dilution or distillation with a catalytic resource. Without the resource,

$$F_{\text{LO}}(|\psi\rangle, |\phi\rangle) = F(P^{\downarrow}, Q^{\downarrow}) \approx 0.958.$$

One can verify that the optimal choice of $r^{\star} \approx 0.6$ in this case. For this choice

$$F_{\text{LU}}(|\psi\rangle|\zeta\rangle, |\phi\rangle|\zeta\rangle) = F((P \otimes R^{\star})^{\downarrow}, (Q \otimes R^{\star})^{\downarrow}) > 0.979.$$

The first problem is that 0.979 is not an acceptably high fidelity even by contemporary standards. Nonetheless, note that to get this state via embezzling (and ignoring that embezzling out the initial state introduces error), it would require generating $|\mu(n)\rangle$ where $n > m^{1/(1-0.979)} = 2 \times 10^{14}$. That is, even to embezzle a two-qubit pure quantum state would require generating an inconceivable amount of entanglement. For this reason, specially engineered embezzling states seem a significant improvement up to any error that can be achieved.

On the other hand, one might note that if we could generate R where $r = 0.55$, then we may as well have just used this state to begin with as

$$F(P^{\downarrow}, R^{\downarrow}) = 0.98989 > F((P \otimes R^{\star})^{\downarrow}, (Q \otimes R^{\star})).$$

From a practical perspective we agree with this critique. Nonetheless, from a basic science perspective, if we are interested in local unitary conversions under embezzling states and catalysts, then the above tells us there are better choices in general than van Dam–Hayden embezzlement, although van Dam–Hayden embezzling has the special property of being universal and optimal for sufficiently small ε.

We close this consideration with two final remarks. First, if one picks two states that are more similar to begin with, then the scaling of the embezzling state will be even larger. Second,

---

[2]We refer the reader to Proposition 3 if the notation has been forgotten.

---

[3]We stress that by the correspondence of Schmidt coefficients to probability distributions as discussed at the start of the work, even without Theorem 1, this would be a legitimate strategy, we simply would not know analytically it was optimal.
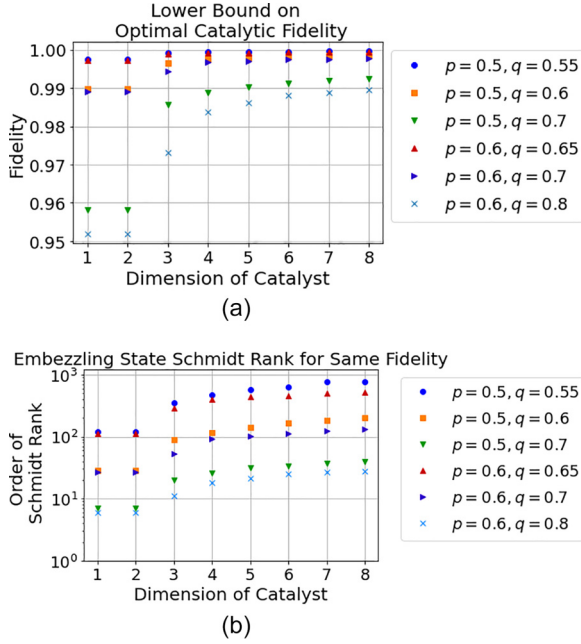
FIG. 5. Plots pertaining to the dimension scaling of embezzling states motivated by Example 5. (a) Depicts a lower bound on the achievable fidelity of converting one two-qubit entangled state to another under local unitaries using an embezzling state with a given local dimension (equivalently, Schmidt rank) where the two-qubit states are parametrized by probability distributions $[p \quad 1-p]$ and $[q \quad 1-q]$. These lower bounds were achieved via brute search as described in the main text. (b) We plot the order (i.e., the power of 10) of the Schmidt rank of the van Dam–Hayden embezzling state $|\mu(n)\rangle$ to obtain the same maximum fidelity. This is calculated using $2^{1/(1-\tilde{F})}$ following Proposition 3. See the main text for further information on this calculation. All chosen embezzling states are provided in Appendix C in Tables I–VI for verification by direct calculation.

we have not presented how the fidelity for this example scales as the local dimension of $|\zeta\rangle$ grows. Both the dimension scaling and two states that are more similar are considered in Fig. 5. There, in Fig. 5(a), we find lower bounds on the optimal fidelity achieved for a given size of embezzling state. To do this, we searched over the discretized (ordered) probability simplex where the discretization was over five-thousandths (i.e., 0.005 intervals in each entry) for dimensions up to 6. For dimensions 7 and 8, due to the time it takes to optimize over such intervals, we would upper bound the search of the $(d+1)$th entry by the dth entry of the previous dimensions optimizer. This heuristic seems natural as the approximate ordered distributions in each dimension decrease entrywise almost all the time (see Appendix C which includes all data for the plot). In some low-dimensional cases, we are able to verify our solutions are near optimal by seeing that the optimizer is achieved over a discretization in only hundredths. To compare to the carefully designed embezzling states of Fig. 5(a), we consider the sufficient Schmidt rank of the van Dam–Hayden universal embezzling family according to Proposition 3. This is done via the following calculation. If a fidelity $\tilde{F}$ is achieved using the specific embezzling state we found, then we are interested in a van Dam–Hayden universal embezzling state $|\mu(n)\rangle$ with Schmidt rank $n > m^{1/\varepsilon}$ such that $1 - \varepsilon \geqslant \tilde{F}$. Thus, we are interested in $n > 2^{1/(1-\tilde{F})}$ where we

have used that in our case the state we are embezzling in and out has Schmidt rank $m = 2$. As one would expect, this is a large number so in Fig. 5(b), we merely plot the *order* (power of 10) of the Schmidt rank. The data for generating this plot is provided in Tables I–IV in Appendix C.

VII. ON EXTENSIONS OF THE THEORY

As a final consideration, we discuss the application of our results beyond bipartite pure quantum states. First we remark upon extensions to multipartite pure quantum states. In this case the problem is that in establishing all of the results, we have used that local unitaries can take the Schmidt decomposition of the state to one of a canonical form. However, in the multipartite case, the Schmidt decomposition does not even exist in general [35]. As such, this argument immediately breaks down. Furthermore, in the proof of Theorem 1 we used Uhlmann's theorem, which requires partitioning the state into two pieces, one of which is the purification. Therefore, it seems no multipartite extension of this work holds.

Similarly, there are issues with approaching mixed states. One issue is to note that all relationships we have been able to establish have stemmed from the fidelity under local unitaries of pure states. Even in the case where local operations made a pure quantum state no longer pure, we purified operations so that the states were pure. We simply cannot do this if we start with mixed states in both arguments of the fidelity. We also cannot purify the states as by data processing, any optimization without tracing off the purifying space only gets us a lower bound. Moreover, this lower bound would require establishing results for tripartite systems, which returns to the issues with the multipartite pure state case. Therefore, we believe in effect these are the most general settings where these proof methods will be of use.

All code used to generate Figs. 4 and 5 may be found at this Git repository [36] for transparency and scrutiny.

ACKNOWLEDGMENTS

APPENDIX A: RANDOMNESS EMBEZZLING PROOF AND DISCUSSION ON LOCALITY

In this Appendix we provide the proof of Proposition 7 and then briefly discuss how it differs from quantum embezzlement.

Proof. The proof is largely the same as for embezzlement of quantum states [14]. Let $P = \sum_i p(i)|i\rangle\langle i|$. Define W_n as $R_n \otimes P$ except with probabilities in decreasing order. Note

$$R_n \otimes P = \frac{1}{H_n} \sum_{i,j} \frac{p(i)}{j} |i\rangle\langle i| \otimes |j\rangle\langle j|,$$

so there exists a relabeling on $\{(i, j)\}$ that will take this to W_n. In particular, letting $f : [m] \times [n] \to [mn]$ be a bijection, we have $|i\rangle|j\rangle \to |f(i, j)\rangle \equiv |i'\rangle|j'\rangle$ such that $\{z_{f(i,j)} := \frac{p(i)}{jH_n}\}_{(i,j)}$ satisfy $z_k \geqslant z_{k+1}$ for all $k \in [mn]$. Therefore, it suffices to

approximate W_n, which means we want to bound the overlap of this with $R_n \otimes P$.

For fixed t and i, we let

$$N_i^t := \left| \left\{ (i, j) : \frac{p(i)}{jH_n} > \frac{1}{tH_n} \right\} \right| .$$

The inequality may be manipulated to imply $1 \leqslant j < p(i)t$. It follows that $N_i^t < p(i)t$. From this we obtain $\sum_{i=1}^{m} N_i^t < \sum_{i=1}^{m} p(i)t < t$, where we have used $\sum_i p(i) = 1$. As $z_1 \geqslant z_2 \geqslant \ldots$, it follows $z_j \leqslant \frac{1}{jH_n}$ for all $1 \geqslant j \geqslant n$. We may restate this as for $1 \leqslant j \leqslant n$, there are at most $t' - 1$ pairs (i, j) such that $p(i)/(jH_n) > 1/(t'H_n)$. Recalling $z_1 \geqslant z_2 \geqslant \ldots$, this means that $z_1 < 1/H_n$ and that there is at most one pair (i, j) pair such that $p(i)/(jH_n) < 1/(2H_n)$, which, since $z_1 \geqslant z_2$, means if such a pair exists, it is z_1. By applying this argument in effect recursively, we see that for t', there are at most $t' - 1$ (i, j) pairs such that $p(i)/(jH_n) > 1/(t'H_n)$ and since $z_k \geqslant z_{k+1}$, if all of these pairs exist, then it must be $z_1, \ldots, z_{t'-1}$. Therefore, $z_j \leqslant 1/(jH_n)$ for all $1 \leqslant j \leqslant n$. We can now use this to bound the fidelity:

$$F(R_n \otimes |0\rangle\langle 0|, W_n) = \left(\sum_{j=1}^{n} \sqrt{\frac{z_j}{jH_n}} \right)^2 \geqslant \left(\sum_{j=1}^{n} \sqrt{z_j} \right)^2$$
$$\geqslant \sum_{j=1}^{n} z_j,$$

where in the equality we have used the definition of fidelity, in the second we used our established inequality, and in the third we have used $\sqrt{x} + \sqrt{y} \geqslant \sqrt{x+y}$ for $x, y \geqslant 0$ to pull the square root out around the sum and cancel with the square.

Now we want to lower bound this sum, which requires managing the z_j terms. We consider $T_n = R_n \otimes \pi_m$ with probabilities $t(j)$ where $\pi_m := \frac{1}{m} \sum_{i=1}^{m} |i\rangle\langle i|$. Now note that $z_k \geqslant t_k$ for all $k \in [mn]$, and this is independent of what the distribution P is. We can then bound the relevant sum by the sum for T_n. It follows

$$\sum_{j=1}^{n} t_j = \sum_{j=1}^{\lfloor n/m \rfloor} \sum_{i=1}^{m} \frac{1}{jH_n m} = \sum_{j=1}^{\lfloor n/m \rfloor} \frac{1}{jH_n} = \frac{H_{\lfloor n/m \rfloor}}{H_m}$$
$$\geqslant \frac{\ln(n/m)}{\ln(n)} = 1 - \frac{\log_2(m)}{\log_2(n)},$$

where the second inequality is using $H_n \geqslant \ln(n)$ and the final form is converting from ln to $\log_2$ in both the numerator and denominator so it cancels. Finally, letting $1 - \log_2(m)/\log_2(n) > 1 - \varepsilon$ will result in $n > m^{1/\varepsilon}$, which completes the proof. ∎

With the proof established, we expand upon the distinction between the entangled and classical distribution cases of embezzlement in terms of locality briefly mentioned in the main text. In the classical case, one party embezzles a distribution locally by themselves, whereas in the entangled case two parties act locally on a nonlocal distribution. Mathematically, this simply follows from the fact the $\text{vec}(\cdot)$ map and its inverse converts between bipartite states and a probability distribution. However, it is also physically interesting that these are the two cases that align as it is clear other

variations are either classically or quantumly impossible as we now explain.

The first reasonable variation would be if there is a nonlocal classical case where two parties try and construct some joint distribution p_{XY} using classical embezzler $r_{X'Y'}$. It is easy to see that they cannot in general satisfy the decoupling condition that is satisfied in quantum embezzlement, i.e., they cannot satisfy $p_{XY} \otimes r_{X'Y'}$ in this setting. This is because without loss of generality the state will be of the form

$$q_{XYX'Y'} = \sum_{x,x',y,y'} q(x|x')q'(y|y')r(x, y)$$
$$\times |x, y, x', y'\rangle\langle x, y, x', y'|.$$

This form means that X will be correlated to X' and Y to Y' unless q_{XY} may be generated nonlocally without a seed state to correlate the two which means they are (up to the allowed error) independent, i.e., $q_{XY} \approx_\varepsilon q_X \otimes q_Y$. In this sense, there cannot be a classical nonlocal equivalent of quantum embezzlement.

On the other hand, if one does not require the decoupling, then this is a task that is possible in the classical setting and is known as distributed source simulation, where the question is the minimal needed *shared randomness* as the *seed state* to generate the target state up to an (arbitrary) error [37]. This was determined asymptotically in the classical case by Wyner [8], extended to separable states by Hayashi [9], and recently generalized to the one-shot setting for separable states in [10]. However, as in this setting variation there is no communication between the acting parties and the embezzler acts as the seed state, it follows from Proposition 2 that distributed source simulation cannot admit an entangled state equivalent. For these reasons, not only does the vec bijection specify the correspondence of embezzlement in the classical and quantum setting, but deviating from it makes either a quantum or classical version impossible.

APPENDIX B: SEMIDEFINITE PROGRAM RELAXATION OF MAX FIDELITY OF PURE STATE TRANSFORMATION UNDER LOSR

In this Appendix we prove Theorem 5. We begin by establishing (11) is true.

Lemma 4. Consider target state $|\psi\rangle$ and seed state $|\phi\rangle$. Let $\text{SR}(\psi) = d$ and $\text{SR}(\phi) = d'$. Define $A = \mathbb{C}^d$, $B = \mathbb{C}^{dd'}$. Then,

$$F_{\text{LOSR}}(|\psi\rangle, |\phi\rangle) \leqslant \max \ F(R, Q_{\text{embed}}^\downarrow) \text{s.t.} \ \text{Tr}_B[R] = P^\downarrow$$
$$\times \ R \in \mathcal{P}^\downarrow(d^2 \cdot d') ,$$

where P and Q are the distributions defined by $|\psi\rangle$ and $|\phi\rangle$'s Schmidt coefficients respectively.

Proof. The above seems intuitively true from Theorem 2 as we have just relaxed the tensor product structure with the partial trace constraint. The technical issue is the ordering operation $\cdot^\downarrow$ is defined in terms of a permutation of a fixed basis, so we need to make sure this works with the partial trace.

Note the feasible set, the set we can optimizer over, in Theorem 2 is $S_1(P) := \{(P \otimes P')^\downarrow : P' \in \mathcal{P}(\Sigma)\}$. Now note

this is the same as the set

$$S_2(P) := \{(P^\downarrow \otimes P'^\downarrow)^\downarrow : P' \in \mathcal{P}(\Sigma)\}$$

because the ordering applied to the tensor product will result in the same thing regardless of whether or not P, P' were ordered. Therefore, we can focus on $P^\downarrow \otimes P'^\downarrow$ to make the explanation clearer.

In general, in terms of vectors,

$$(p^\downarrow \otimes p'^\downarrow)^\downarrow = \begin{pmatrix} p^\downarrow(1)p'^\downarrow \\ p^\downarrow(2)p'^\downarrow \\ \vdots \\ p^\downarrow(d)p'^\downarrow \end{pmatrix},$$

where $p(i) \geqslant p(i+k)$ for $k \geqslant 0$. Formally, we also have

$$p^\downarrow(i)p'^\downarrow(1) \geqslant p^\downarrow(i+k)p'^\downarrow(j)$$

for all $i \in [d]$, $k \in \{0, \ldots, d-i\}$, and $j \in \Sigma$. In particular, what this means is that without loss of generality for any $i \in [d]$, $p^\downarrow(i)p'^\downarrow(1)$ appears before any element that is *not* of the form $p^\downarrow(i-\ell)p'^\downarrow(j)$ for some $0 < \ell \leqslant i-1$. It follows that under the ordering of $(p^\downarrow \otimes p'^\downarrow)^\downarrow$, when the partial trace marginalizes to the A space, the induced ordering on the local space will be the ordering based on $p^\downarrow$. Formally, this can be expressed as

$$\mathrm{Tr}_{\mathbb{C}^{|\Sigma|}}[(P^\downarrow \otimes P'^\downarrow)^\downarrow] = \sum_{j \in \Sigma} \mathbb{1}_A \otimes \langle j|(P^\downarrow \otimes P'^\downarrow)^\downarrow|j\rangle$$

$$= \sum_{i \in [d]} p^\downarrow(i)|i\rangle\langle i|,$$

where the first equality is a representation of the partial trace and the second is using the property noted of the ordering on the joint ordered distribution.

Thus, if $X \in S_2(P)$, $\mathrm{Tr}_{C^{|\Sigma|}}(X) = P^\downarrow$ and $X \in \mathcal{P}^\downarrow(d|\Sigma|)$. Noting that $|\Sigma| = dd'$, this is the feasible set we have defined in the proposition. This completes the proof. ∎

The remaining point is to prove this is the semidefinite program given in (12). There is much to the theory of semidefinite programs for quantum information [27], but for our purposes all we will need is the following definition.

Definition 4. A semidefinite program may be expressed as

$$\max \mathrm{Tr}(AX)\,\text{s.t.}\ \Phi(X) = BX_{\mathbb{C}^d} \preceq 0,$$

where $\Phi \in \mathrm{T}(\mathbb{C}^d, \mathbb{C}^{d'})$ is a Hermitian-preserving map, $A \in \mathrm{Herm}(\mathbb{C}^d)$, $B \in \mathrm{Herm}(\mathbb{C}^{d'})$, and $\mathrm{Herm}(\cdot)$ is the space of Hermitian operators on a given Hilbert space.

The fidelity is known to be a semidefinite program [27], so we are really just verifying all of our constraints work and that we can write the SDP simply by making use of that.

Lemma 5. The optimization program in the previous lemma may be expressed as the following semidefinite program over the reals:

$$\max \sum_{i \in [d^2 d']} x(i)\,\text{s.t.}\ \begin{pmatrix} \mathrm{diag}(r)\mathrm{diag}(x) \\ \mathrm{diag}(x) & \mathrm{diag}(q^\downarrow_{\mathrm{embed}}) \end{pmatrix} \succeq 0$$

$$\times\ \mathrm{Tr}_B[\mathrm{diag}(r)] = P^\downarrow r \in \mathcal{P}^\downarrow([d^2 d])x \in \mathbb{R}^{d^2 d'},$$

where d, d' are defined in the previous lemma.

Proof. We begin by expressing the objective function of the previous lemma, which is in terms of fidelity, using the primal problem for the SDP for fidelity from [27, Theorem 3.17]:

$$\max \frac{1}{2}[\mathrm{Tr}(X) + \mathrm{Tr}(X^\dagger)] \begin{pmatrix} R & X \\ X^\dagger & Q^\downarrow_{\mathrm{embed}} \end{pmatrix}$$

$$\geqslant 0X \in \mathrm{L}(\mathbb{C}^{[d^2 \cdot d']}).$$

Now our goal is to reduce X to the diagonal of a real vector.

Note that $R, Q^\downarrow_{\mathrm{embed}}$ are always invariant under pinching onto the computational basis of $\mathbb{C}^{[d^2 \cdot d']}$, which we can denote Δ. Note that this pinching is a CPTP, so by the CP property,

$$(\mathrm{id}_{\mathbb{C}^2} \otimes \Delta)\begin{pmatrix} R & X \\ X^\dagger & Q^\downarrow_{\mathrm{embed}} \end{pmatrix} = \begin{pmatrix} R & \Delta(X) \\ \Delta(X^\dagger) & Q^\downarrow_{\mathrm{embed}} \end{pmatrix} \geqslant 0.$$

It also then follows as a positive-semidefinite operator is always Hermitian that

$$\begin{pmatrix} R & \Delta(X^\dagger) \\ \Delta(X) & Q^\downarrow_{\mathrm{embed}} \end{pmatrix} \geqslant 0.$$

Thus, by taking these two cases and averaging them, we have that

$$\begin{pmatrix} R & \frac{1}{2}(\Delta(X + X^\dagger)) \\ \frac{1}{2}(\Delta(X + X^\dagger)) & Q^\downarrow_{\mathrm{embed}} \end{pmatrix} \geqslant 0.$$

Define $\overline{X} := \frac{1}{2}(\Delta(X + X^\dagger))$. Then note

$$\frac{1}{2}(\mathrm{Tr}(X) + \mathrm{Tr}(X^\dagger)) = \frac{1}{2}(\mathrm{Tr}[\Delta(X)] + \mathrm{Tr}[\Delta(X^\dagger)])$$

$$= \frac{1}{2}(\mathrm{Tr}(\overline{X}) + \mathrm{Tr}(\overline{X}^\dagger)) = \mathrm{Tr}(\overline{X}),$$

where the first equality is because the pinching is trace preserving, the second is by definition of $\overline{X}$, as is the final equality. Thus, for any X that satisfies the positivity constraint, we could replace it with $\overline{X}$ without loss of generality as we are considering a maximization. Finally, note that $\overline{X}$ is a real diagonal matrix by the pinching along with the fact $a + a^* = 2\,\mathrm{Re}a$. Thus, $\overline{X} = \mathrm{diag}(x)$ for some $x \in \mathbb{R}^{d^2 d'}$. Combining all these points and using $\mathrm{Tr}(\overline{X}) = \sum_{i \in [d^2 d']} x(i)$, we have reduced to considering

$$\max \sum_{i \in [d^2 d']} x(i) \begin{pmatrix} \mathrm{diag}(r) & \mathrm{diag}(x) \\ \mathrm{diag}(x) & \mathrm{diag}(q^\downarrow_{\mathrm{embed}}) \end{pmatrix} \geqslant 0x \in \mathbb{R}^{d^2 d'}.$$

This argument works for any choice of diagonal r, so this is the major reduction.

What remains is to prove all the constraints are Hermitian maps. One can write the constraints for $r \in \mathcal{P}^\downarrow$ as $r(i) \geqslant r(i+1)$ for all i, which are semidefinite constraints and can be written as Hermitian-preserving maps on the variables r, x. diag is a Hermitian-preserving map as is the partial trace, so $\mathrm{Tr}_C[\mathrm{diag}(r)]$ is a Hermitian-preserving map. Likewise is the block-matrix mapping if one allows for the complex conjugate in the lower left block, but noting $\mathrm{diag}(x)^\dagger = \mathrm{diag}(x)$, we can leave it as written. Thus, all the maps are Hermitian preserving.

The conversion to actual standard form we then omit as it provides no insight. This completes the proof. ∎

The above two proofs establish Theorem 5.

APPENDIX C: DATA FOR EMBEZZLING STATE FIGURE

In this Appendix, we provide all the embezzling distributions used to generate Fig. 5.

TABLE I. Data for the near-optimal catalyst results presented in Fig. 5 in the case $p = 0.5$ and $q = 0.55$. d stands for the dimension of the distribution. Each distribution r was found via numerical search as described in the main text.

d	Lower bound embezzling distribution r
1	n/a
2	[0.5,0.5]
3	[0.3,0.33,0.37]
4	[0.215,0.235,0.26,0.29]
5	[0.16,0.18,0.2,0.22,0.24]
6	[0.13,0.145,0.155,0.17,0.19,0.21]
7	[0.105,0.115,0.125,0.14,0.155,0.17,0.19]
8	[0.085,0.095,0.105,0.115,0.13,0.115,0.13,0.145,0.155,0.17]

TABLE II. Data for the near-optimal catalyst results presented in Fig. 5 in the case $p = 0.5$ and $q = 0.6$. d stands for the dimension of the distribution. Each distribution r was found via numerical search as described in the main text.

d	Lower bound embezzling distribution r
1	n/a
2	[0.5,0.5]
3	[0.27,0.33,0.4]
4	[0.18,0.22,0.27,0.33]
5	[0.13,0.16,0.195,0.235,0.28]
6	[0.095,0.115,0.14,0.175,0.215,0.26]
7	[0.075,0.09,0.11,0.13,0.16,0.195,0.24]
8	[0.055,0.065,0.08,0.1,0.125,0.155,0.19,0.23]

TABLE III. Data for the near-optimal catalyst results presented in Fig. 5 in the case $p = 0.5$ and $q = 0.7$. d stands for the dimension of the distribution. Each distribution r was found via numerical search as described in the main text.

d	Lower bound embezzling distribution r
1	n/a
2	[0.5,0.5]
3	[0.21,0.32,0.47]
4	[0.12,0.185,0.28,0.415]
5	[0.075,0.115,0.175,0.26,0.375]
6	[0.05,0.075,0.11,0.165,0.245,0.355]
7	[0.03,0.045,0.07,0.105,0.16,0.24,0.35]
8	[0.05,0.065,0.085,0.095,0.13,0.14,0.19,0.245]

TABLE IV. Data for the near-optimal catalyst results presented in Fig. 5 in the case $p = 0.6$ and $q = 0.65$. d stands for the dimension of the distribution. Each distribution r was found via numerical search as described in the main text.

d	Lower bound embezzling distribution r
1	n/a
2	[0.5,0.5]
3	[0.19,0.31,0.5]
4	[0.1,0.165,0.275,0.46]
5	[0.055,0.095,0.155,0.26,0.435]
6	[0.07,0.11,0.12,0.185,0.2,0.315]
7	[0.04,0.065,0.105,0.115,0.175,0.19,0.31]
8	[0.035,0.06,0.065,0.1,0.11,0.165,0.18,0.285]

TABLE V. Data for the near-optimal catalyst results presented in Fig. 5 in the case $p = 0.6$ and $q = 0.7$. d stands for the dimension of the distribution. Each distribution r was found via numerical search as described in the main text.

d	Lower bound embezzling distribution r
1	n/a
2	[0.5,0.5]
3	[0.24,0.275,0.485]
4	[0.08,0.15,0.27,0.5]
5	[0.04,0.075,0.14,0.26,0.485]
6	[0.055,0.095,0.11,0.18,0.205,0.355]
7	[0.03,0.055,0.09,0.11,0.17,0.2,0.345
8	[0.03,0.05,0.055,0.09,0.1,0.165,0.185,0.325]

TABLE VI. Data for the near-optimal catalyst results presented in Fig. 5 in the case $p = 0.6$ and $q = 0.8$. d stands for the dimension of the distribution. Each distribution r was found via numerical search as described in the main text.

d	Lower bound embezzling distribution r
1	n/a
2	[0.5,0.5]
3	[0.21,0.29,0.5]
4	[0.09,0.175,0.24,0.495]
5	[0.05,0.095,0.16,0.23,0.465]
6	[0.035,0.065,0.09,0.155,0.215,0.44]
7	[0.02,0.04,0.065,0.09,0.15,0.21,0.425]
8	[0.015,0.03,0.04,0.065,0.09,0.145,0.205,0.41]

[1] J. A. Wheeler, Information, physics, quantum: The search for links, in *Proceedings III International Symposium on Foundations of Quantum Mechanics* (Physical Society of Japan, Tokyo, 1989), pp. 354–358.

[2] *Complexity, Entropy, and the Physics of Information*, Vol. VIII, edited by W. H. Zurek (Westview Press, Boulder, CO, 1989).

[3] R. Landauer, Information is physical, Phys. Today **44**(5), 23 (1991).

[4] E. Chitambar and G. Gour, Quantum resource theories, Rev. Mod. Phys. **91**, 025001 (2019).

[5] J. S. Bell, On the Einstein Podolsky Rosen paradox, Phys. Phys. Fiz. **1**, 195 (1964).

[6] P. Hayden and A. Winter, Communication cost of entanglement transformations, Phys. Rev. A **67**, 012326 (2003).

[7] A. Harrow and H.-K. Lo, A tight lower bound on the classical communication cost of entanglement dilution, IEEE Trans. Inf. Theory **50**, 319 (2004).

[8] A. Wyner, The common information of two dependent random variables, IEEE Trans. Inf. Theory **21**, 163 (1975).

[9] M. Hayashi, *Quantum Information: An Introduction* (Springer, Berlin, 2006).

[10] I. George, M.-H. Hsieh, and E. Chitambar, One-shot distributed source simulation: As quantum as it can get, arXiv:2301.04301.

[11] D. Schmid, H. Du, M. Mudassar, G. Coulter-de Wit, D. Rosset, and M. J. Hoban, Postquantum common-cause channels: The resource theory of local operations and shared entanglement, Quantum **5**, 419 (2021).

[12] C. Datta, T. V. Kondra, M. Miller, and A. Streltsov, Catalysis of entanglement and other quantum resources, Rep. Prog. Phys. **86**, 116002 (2023).

[13] P. Lipka-Bartosik, H. Wilming, and N. H. Y. Ng, Catalysis in quantum information theory arXiv:2306.00798.

[14] W. van Dam and P. Hayden, Universal entanglement transformations without communication, Phys. Rev. A **67**, 060302(R) (2003).

[15] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, The quantum reverse shannon theorem and resource tradeoffs for simulating quantum channels, IEEE Trans. Inf. Theory **60**, 2926 (2014).

[16] A. Anshu, S. B. Hadiashar, R. Jain, A. Nayak, and D. Touchette, One-shot quantum state redistribution and quantum markov chains, in *2021 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Piscataway, NJ, 2021), pp. 130–135.

[17] D. Leung and B. Wang, Characteristics of universal embezzling families, Phys. Rev. A **90**, 042331 (2014).

[18] I. Dinur, D. Steurer, and T. Vidick, A parallel repetition theorem for entangled projection games, Comput. Complexity **24**, 201 (2015).

[19] P. Hayden, R. Jozsa, D. Petz, and A. Winter, Structure of states which satisfy strong subadditivity of quantum entropy with equality, Commun. Math. Phys. **246**, 359 (2004).

[20] G. Gour, Resources of the quantum world, arXiv:2402.05474.

[21] F. Buscemi, All entangled quantum states are nonlocal, Phys. Rev. Lett. **108**, 200401 (2012).

[22] G. Aubrun and I. Nechita, Catalytic majorization and norms, Commun. Math. Phys. **278**, 133 (2008).

[23] D. Leung, B. Toner, and J. Watrous, Coherent state exchange in multi-prover quantum interactive proof systems, Chicago J. Theor. Comput. Sci. **11**, 1 (2013).

[24] F. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, The second laws of quantum thermodynamics, Proc. Natl. Acad. Sci. USA **112**, 3275 (2015).

[25] N. H. Y. Ng, L. Mančinska, C. Cirstoiu, J. Eisert, and S. Wehner, Limits to catalysis in quantum thermodynamics, New J. Phys. **17**, 085004 (2015).

[26] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, 2013).

[27] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, 2018).

[28] J. I. de Vicente and M. Huber, Multipartite entanglement detection from correlation tensors, Phys. Rev. A **84**, 062306 (2011).

[29] G. S. Mudholkar and M. Freimer, A structure theorem for the polars of unitarily invariant norms, Proc. Am. Math. Soc. **95**, 331 (1985).

[30] N. Johnston, Norms and cones in the theory of quantum entanglement, Ph.D. thesis, University of Guelph, 2012.

[31] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner, The uncertainty principle in the presence of quantum memory, Nat. Phys. **6**, 659 (2010).

[32] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. (Cambridge University Press, Cambridge, 2013).

[33] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Concentrating partial entanglement by local operations, Phys. Rev. A **53**, 2046 (1996).

[34] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, Hoboken, NJ, 2006).

[35] A. Peres, Higher order schmidt decompositions, Phys. Lett. A **202**, 16 (1995).

[36] https://github.com/ChitambarLab/ RevisitingPureStateTransformationswithZeroCommunication.

[37] L. Yu and V. Y. F. Tan, Common information, noise stability, and their extensions, Found. Trends Commun. Inf. Theory **19**, 107 (2022).