Optimizing practical entanglement distillation

Rozpdek, Filip; Schiet, Thomas; Thinh, Le Phuc; Elkouss, David; Doherty, Andrew C.; Wehner, Stephanie

DOI
10.1103/PhysRevA.97.062333

Publication date
2018

Document Version
Final published version

Published in
Physical Review A

Optimizing practical entanglement distillation

Filip Rozpędek,[1,*] Thomas Schiet,[1] Le Phuc Thinh,[1] David Elkouss,[1] Andrew C. Doherty,[2] and Stephanie Wehner[1]

[1] *QuTech, Lorentzweg 1, 2628 CJ Delft, Netherlands*

[2] *Centre for Engineered Quantum Systems, School of Physics, University of Sydney, Sydney, NSW 2006, Australia*

(Received 6 April 2018; published 21 June 2018)

The goal of entanglement distillation is to turn a large number of weakly entangled states into a smaller number of highly entangled ones. Practical entanglement distillation schemes offer a trade-off between the fidelity to the target state and the probability of successful distillation. Exploiting such trade-offs is of interest in the design of quantum repeater protocols. Here, we present a number of methods to assess and optimize entanglement distillation schemes. We start by giving a numerical method to compute upper bounds on the maximum achievable fidelity for a desired probability of success. We show that this method performs well for many known examples by comparing it to well-known distillation protocols. This allows us to show optimality for many well-known distillation protocols for specific states of interest. As an example, we analytically prove optimality of the distillation protocol utilized within the Extreme Photon Loss entanglement generation scheme, even in the asymptotic limit. We proceed to present a numerical method that can improve an existing distillation scheme for a given input state, and we present an example for which this method finds an optimal distillation protocol. An implementation of our numerical methods is available as a Julia package.

DOI: 10.1103/PhysRevA.97.062333

I. INTRODUCTION

Entanglement distillation forms an important element of many proposals for quantum repeaters [1–5], as well as networked quantum computers [6,7]. It has seen widespread study across several areas ranging from practical entanglement distillation schemes [7–13] and their experimental implementations [14–18], to a general understanding of some of its possibilities and limitations in quantum information theory [19]. The general goal of bipartite entanglement distillation is to convert a state ρ_{AB} into a state $\eta_{\hat{A}\hat{B}}$ that is close to a maximally entangled state $\Phi_{\hat{A}\hat{B}}$ using only local operations and classical communication (LOCC) between the network node holding A (Alice) and the one holding B (Bob). Here by A and B we denote the input registers and by $\hat{A}$ and $\hat{B}$ the output ones. Closeness is measured in terms of the fidelity

$$F = \langle \Phi_D | \eta_{\hat{A}\hat{B}} | \Phi_D \rangle \geqslant 1 - \epsilon \tag{1}$$

to the target state

$$|\Phi_D\rangle = \frac{1}{\sqrt{D}} \sum_{j=0}^{D-1} |j\rangle_{\hat{A}} |j\rangle_{\hat{B}}, \tag{2}$$

which is maximally entangled across $\hat{A}$ and $\hat{B}$.

There is a slight difference between the meaning of *entanglement distillation* in the quantum information theory literature and in practical schemes. In quantum information theory, one typically considers the case where $\rho_{AB} \approx (\tau_{ab})^{\otimes n}$ consist of n copies of a state τ_{ab}. If we want to distill states that are arbitrarily close to the perfect maximally entangled state, then the distillable entanglement $E_D(\tau_{ab})$ of τ_{ab} answers the question of how large this output state can be. Specifically, it tells us what would be the dimension $|\hat{A}\hat{B}|$ relative to the input dimension $|AB|$, under distillation using LOCC as $n \to \infty$ [20]. As such, the dimension of the output state $|\hat{A}\hat{B}|$ is generally smaller than the dimension $|AB|$ of the input state, unless the input is already maximally entangled. While E_D is difficult to compute in general, several computable bounds have been proposed [21–24]. Recent years have seen one-shot variants of distillable entanglement in which n can be finite, or indeed ρ_{AB} may have an arbitrary structureless form [25–27]. Bounds on the one-shot distillable entanglement may be computed numerically [28]. Crucially, the task of entanglement distillation as it is considered in quantum information theory always produces an output state $\eta_{\hat{A}\hat{B}}$, and considers no failure. The possibility of failure is allowed implicitly by assuming that if the entanglement distillation procedure fails, then Alice and Bob output an arbitrary state leading to a reduced fidelity of the output state to the target state.

In contrast, practical schemes for entanglement distillation explicitly allow for the possibility of failure [7–13]. The fidelity F to the target state is in that case of interest only in the event of success. Not surprisingly, there exist interesting trade-offs between this fidelity F and the probability of success p_{succ} of the distillation procedure. A simple example of such a trade-off is the possibility of *filtering* in which the dimensions $|\hat{A}|$ and $|\hat{B}|$ of the output systems $\hat{A}$ and $\hat{B}$ are equal to the input dimensions $|A|$ and $|B|$, that is, $|\hat{A}| = |A|$ and $|\hat{B}| = |B|$. Yet, it is possible to probabilistically increase the fidelity to the target state by LOCC, where a higher fidelity F leads to a lower success probability p_{succ}. More generally, trading off the fidelity F against p_{succ} is relevant to the construction of quantum networks: here, the initial generation of entanglement is typically already probabilistic such as when using a heralded scheme to produce the initial (imperfect) entanglement [29,30].

[*] f.d.rozpedek@tudelft.nl

Most significantly, however, the local quantum memory used to store entanglement is itself imperfect. This means that both the initial as well as the resulting entanglement cannot be preserved for an arbitrary amount of time. Clearly, the success probability p_{succ} dictates the rate at which we can hope to produce high-fidelity entanglement between different nodes in the network. This rate imposes requirements on the coherence times of the memory if multiple entangled pairs are generated such that they should undergo further processing, for example, to generate more complex entangled states in a multinode network. In such a scenario, one may thus wish to obtain a higher probability of success at the expense of a lower fidelity (or vice versa) in relation to the local storage capabilities of the nodes.

Due to a limited lifetime of local quantum memories, practical distillation schemes are not expected to employ multiround operations in the near future. Instead, practically employed schemes consist of applying a local operation and measurement on Alice's and Bob's side, followed by a single exchange of measurement outcomes using classical communication in order to decide success or failure. Here, we will refer to this subset of LOCC as *measure and exchange (MX) operations* due to their reduced technical demands (see Sec. III A for a definition).

II. OVERVIEW

In this paper, we develop *a set of tools* for optimizing and assessing existing practical distillation schemes. Specifically, our tools allow for a detailed investigation of the trade-off between the possible output fidelity and probability of success of distillation schemes.

In Sec. III A, we first formally define the set of measure and exchange (MX) operations, and illustrate it with an example of an existing filtering protocol.

In Sec. III C, we state a semidefinite programming (SDP) method to compute upper bounds on the achievable fidelity (or success probability) of a distillation scheme for a given success probability (or fidelity). These methods adapt the ideas of Rains [21] as well as the later methods of Bose symmetric extensions [31,32] to the case of MX operations, where immediate measurements are performed to decide success or failure. We implement these methods in a numerical package that is freely available on GitHub [33].

In Sec. III D, we present a numerical seesaw method based on semidefinite programming that takes a specific distillation scheme and entangled state as input, and iteratively searches for a better distillation scheme adapted to the state of interest. This method is also included in our numerical package.

In Sec. IV, we illustrate our method with a variety of examples, considering different entangled states of interest. We compare upper bounds attained with existing distillation schemes (and interpolations between existing distillation schemes) to determine their performance. We observe optimality for a number of schemes for specific states of interest, including modifications of such schemes and certain new schemes obtained from existing ones using our tools. Specifically, we present an instance in which the seesaw method will find an optimal distillation scheme from an existing one that is suboptimal for the given state.

In the Appendices (summary in Sec. IV) we employ our semidefinite programming methods to analytically prove optimality of the DEJMPS protocol [9] for distilling Bell diagonal states of rank up to three. Furthermore we show optimality of the distillation procedure used within the Extreme Photon Loss (EPL) remote entanglement generation scheme as described in Refs. [7,13], even in the limit of asymptotically many copies.

III. OPTIMIZATION METHODS

Let us now first define MX operations, and specify the problem of interest in terms of such operations. Throughout, we will use the convention $\sigma_X = \text{tr}_Y(\sigma_{XY})$ to denote the marginal σ_X of a larger state σ_{XY}. Moreover, for the purpose of the compactness of notation, we will often omit writing explicitly the identity matrix or the identity channel. That is, for $(\mathbb{I}_A \otimes M_B)\rho_{AB}$ we will often use the shorthand $M_B\rho_{AB}$ and for $(\mathbb{1}_A \otimes \Lambda_{B\to\hat{B}})(\rho_{AB})$ we will use $\Lambda_{B\to\hat{B}}(\rho_{AB})$.

A. Measure and exchange (MX) operations

All MX operations can be modeled as completely positive trace-preserving (CPTP) maps; e.g., for Alice

$$\Lambda_{A\to\hat{A}F_A} : \mathcal{D}(\mathcal{H}_A) \to \mathcal{D}(\mathcal{H}_{\hat{A}F_A}), \tag{3}$$

where $\mathcal{H}_A$ and $\mathcal{H}_{\hat{A}F_A} := \mathcal{H}_{\hat{A}} \otimes \mathcal{H}_{F_A}$ denote the input and output spaces, respectively, and $\mathcal{D}$ denotes the set of density operators living on the space. The registers F_A and F_B denote classical flag registers, which Alice and Bob will compare in order to decide success or failure. Applying these maps locally yields the state

$$\sigma_{\hat{A}F_A\hat{B}F_B} = \Lambda_{A\to\hat{A}F_A} \otimes \Lambda_{B\to\hat{B}F_B}(\rho_{AB}). \tag{4}$$

Since Alice and Bob use classical communication to compare the flags, we may without loss of generality assume that the state after a measurement on F_A and F_B is of the form

$$\sigma_{\hat{A}\hat{B}F_AF_B} = \sum_{f_A,f_B} \sigma_{\hat{A}\hat{B}}^{f_A,f_B} \otimes |f_A\rangle\langle f_A|_{F_A} \otimes |f_B\rangle\langle f_B|_{F_B}, \tag{5}$$

where the sum is taken over strings f_A and f_B, and $0 \leqslant \text{tr}(\sigma_{\hat{A}\hat{B}}^{f_A,f_B}) \leqslant 1$. Comparing the flags to decide success or failure can be understood as subsequently projecting the state using a projector

$$P_{\checkmark} = \sum_{(f_A,f_B)\in\mathcal{S}} |f_A\rangle\langle f_A|_{F_A} \otimes |f_B\rangle\langle f_B|_{F_B}, \tag{6}$$

where $\mathcal{S} = \{(f_A,f_B) \mid \text{Alice and Bob declare success}\}$. The success probability can thus be expressed as

$$p_{\text{succ}} = \text{tr}\left(P_{\checkmark}\sigma_{F_AF_B}\right). \tag{7}$$

The global state conditioned on success can in turn be written as

$$\eta_{\hat{A}\hat{B}F_AF_B} = \frac{(\mathbb{I}_{\hat{A}\hat{B}} \otimes P_{\checkmark})\sigma_{\hat{A}\hat{B}F_AF_B}(\mathbb{I}_{\hat{A}\hat{B}} \otimes P_{\checkmark})}{p_{\text{succ}}}, \tag{8}$$

which has a fidelity to the ideal maximally entangled state

$$F = \langle\Phi_D|\eta_{\hat{A}\hat{B}}|\Phi_D\rangle. \tag{9}$$

Our formalism captures all practical schemes by appropriate definition of $P_{\checkmark}$.

As an example let us consider the filtering protocol [34]. This protocol is adapted to perform well for an input state with $|A| = |B| = 2$ of the form

$$\rho_{AB} = p|\Phi_2\rangle\langle\Phi_2| + (1-p)|01\rangle\langle 01|. \tag{10}$$

In this procedure, Alice performs a measurement given by the POVM: $\{M_A^0, M_A^1\}$ with $M_A^1 = (A_A^1)^\dagger A_A^1$, where $A_A^1 = \sqrt{\epsilon}|0\rangle\langle 0| + |1\rangle\langle 1|$ and $M_A^0 = (A_A^0)^\dagger A_A^0 = \mathbb{I} - M_A^1$ for some parameter ϵ determining the trade-off between F and p_{succ}. In terms of the map this measurement can be expressed as

$$\Lambda_{A\to\hat{A},F_A}(\rho) = \sum_{f_A\in\{0,1\}} A_A^{f_A}\rho\big(A_A^{f_A}\big)^\dagger \otimes |f_A\rangle\langle f_A|_{F_A}. \tag{11}$$

Similarly, Bob performs a measurement given by the POVM: $\{M_B^0, M_B^1\}$ with $M_B^1 = (A_B^1)^\dagger A_B^1$, where $A_B^1 = \sqrt{\epsilon}|1\rangle\langle 1| + |0\rangle\langle 0|$ and $M_B^0 = (A_B^0)^\dagger A_B^0 = \mathbb{I} - M_B^1$, giving the map

$$\Lambda_{B\to\hat{B},F_B}(\rho) = \sum_{f_B\in\{0,1\}} A_B^{f_B}\rho\big(A_B^{f_B}\big)^\dagger \otimes |f_B\rangle\langle f_B|_{F_B}. \tag{12}$$

Alice and Bob declare success if $f_A = f_B = 1$, corresponding to a choice of $P_{\checkmark} = |11\rangle\langle 11|_{F_AF_B}$.

When optimizing over measure and exchange operations, it is sometimes convenient to consider a slightly more general class of operations which we call *measure and exchange operations with shared randomness (MXS operations)*. As the name suggests, Alice and Bob have additional access to classical shared randomness, which is easy to distribute ahead of time. Specifically, if Alice and Bob have a classical symbol r chosen with probability p_r, then they can perform MX operations that depend on r. This means the output state is of the form

$$\sigma_{\hat{A}\hat{B}F_AF_B} = \sum_r p_r \Lambda_{r,A\to\hat{A}F_A} \otimes \Lambda_{r,B\to\hat{B}F_B}(\rho_{AB}). \tag{13}$$

Note the set of MXS operations is a convex set unlike the set of MX operations.

B. Optimizing over MX operations

1. General form

We are now going to consider various optimizations related to the distillation problem. As we have seen, we would like to optimize one of the three parameters D, p_{succ}, ϵ, where D is the local output dimension, p_{succ} is the success probability, and the fidelity is $1-\epsilon$. We will typically fix the output dimension D and for now we will consider optimizing the fidelity for fixed success probability $p_{\text{succ}} = \delta$. It is straightforward to adapt the techniques below to optimize p_{succ} instead. Ideally, we thus wish to solve the following (quadratic) optimization problem over maps $\Lambda_{A\to\hat{A}F_A}$ and $\Lambda_{B\to\hat{B}F_B}$:

$$\begin{aligned}
\text{maximize} \quad & \frac{1}{\delta}\operatorname{tr}\left(|\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}} \otimes P_{\checkmark}\, \sigma_{\hat{A}\hat{B}F_AF_B}\right)\\
\text{subject to} \quad & \operatorname{tr}\left(P_{\checkmark}\sigma_{F_AF_B}\right) = \delta,\\
& \sigma_{\hat{A}\hat{B}F_AF_B} = \Lambda_{A\to\hat{A}F_A} \otimes \Lambda_{B\to\hat{B}F_B}(\rho_{AB}).
\end{aligned}$$

Optimization Program 1.

2. Simplifying the optimization problem

How do we optimize over quantum operations? The key is to employ the Choi isomorphism which gives a one-to-one correspondence between quantum channels and quantum states with certain properties. Specifically, for any quantum channel $\Gamma_{S\to R}$ from a system S to system R, there corresponds a unique Choi state

$$C_{RS'} = \Gamma_{S\to R} \otimes \mathbb{1}_{S'}(\Phi_{SS'}), \tag{14}$$

satisfying

$$C_{RS'} \geqslant 0, \quad C_{S'} = \frac{\mathbb{I}_{S'}}{|S|}, \tag{15}$$

where $\Phi_{SS'}$ is the density matrix of the normalized maximally entangled state from Eq. (2) of dimension $D = |S|$. The Choi state carries all information of the original channel, in the sense that

$$\operatorname{tr}[M_R\Gamma_{S\to R}(\rho_S)] = |S|\operatorname{tr}\left[M_R \otimes \rho_{S'}^T(C_{RS'})\right] \tag{16}$$

for all matrices M_R on R.

For the case of MX operations the Choi states take a product form. This is because a maximally entangled state of a larger system whose dimension D is a composite number is formed by taking the tensor product of maximally entangled states:

$$\begin{aligned}
C_{\hat{A}F_A\hat{B}F_B,A'B'} &= \Lambda_{A\to\hat{A}F_A} \otimes \Lambda_{B\to\hat{B}F_B}(\Phi_{AA'} \otimes \Phi_{BB'})\\
&= C_{\hat{A}F_AA'} \otimes C_{\hat{B}F_BB'}.
\end{aligned} \tag{17}$$

This translates the optimization to the space of product of two Choi states. Similarly, for MXS operations we obtain the optimization over the subset of separable Choi states that can be decomposed as follows (we denote this set here as SEP-C):

$$C_{\hat{A}F_A\hat{B}F_B,A'B'} = \sum_r p_r C_{r,\hat{A}F_AA'} \otimes C_{r,\hat{B}F_BB'}. \tag{18}$$

Note that SEP-C is a strict subset of the set SEP of separable states, since we require that the individual components satisfy the Choi condition Eq. (15).

Before delving into the various approaches to optimize our function below, let us first simplify the problem slightly. Our goal will be to remove the registers F_A and F_B from the expressions above. In particular, let us imagine that $C^*_{\hat{A}F_A,A'}$ and $C^*_{\hat{B}F_B,B'}$ are optimal solutions to the optimization problem above. We then claim that

$$\tilde{C}_{\hat{A}F_A,A'} = \sum_{f_A\in\{0,1\}} |f_A\rangle\langle f_A|_{F_A} C^*_{\hat{A}F_AA'} |f_A\rangle\langle f_A|_{F_A}, \tag{19}$$

$$\tilde{C}_{\hat{B}F_B,B'} = \sum_{f_B\in\{0,1\}} |f_B\rangle\langle f_B|_{F_B} C^*_{\hat{B}F_BB'} |f_B\rangle\langle f_B|_{F_B} \tag{20}$$

are also optimal. This is an immediate consequence of the fact that in our optimization problem, we always measure the registers F_A and F_B. We can thus without loss of generality assume that both states are cq states:

$$\tilde{C}_{\hat{A}F_AA'} = \sum_{f_A\in\{0,1\}} \hat{C}_{f_A,\hat{A}A'} \otimes |f_A\rangle\langle f_A|_{F_A}, \tag{21}$$

$$\tilde{C}_{\hat{B}F_BB'} = \sum_{f_B\in\{0,1\}} \hat{C}_{f_B,\hat{B}B'} \otimes |f_B\rangle\langle f_B|_{F_B}; \tag{22}$$

that is the flags are always classical registers.

Observing that our optimization problem is only concerned with the case that Alice and Bob succeed, we can now express the problem in terms of the Choi states. We can now consider two cases:

(1) Some protocols have local success flags; e.g., the protocol succeeds if Alice and Bob both measure "1", which is the case in the filtering protocol described in Sec. III A or the distillation protocol used within the EPL scheme (both are also described in Appendix B 1). The meaning of "local" refers to the fact that here Alice and Bob can individually already declare failure if they observe a "0" (success evidently requires a comparison). For this example we arrive at the optimization problem

$$\begin{aligned}\text{maximize}\quad & \frac{|A||B|}{\delta}\,\mathrm{tr}\left[|\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}}\otimes\rho^T_{A'B'}(\hat{C}_{1,\hat{A}A'}\otimes\hat{C}_{1,\hat{B}B'})\right]\\ \text{subject to}\quad & |A||B|\,\mathrm{tr}\left[\rho^T_{A'B'}(\hat{C}_{1,A'}\otimes\hat{C}_{1,B'})\right]=\delta,\\ & \hat{C}_{1,\hat{A}A'}\geqslant 0,\quad \hat{C}_{1,\hat{B}B'}\geqslant 0,\\ & \hat{C}_{1,A'}\leqslant\frac{\mathbb{I}_{A'}}{|A|},\quad \hat{C}_{1,B'}\leqslant\frac{\mathbb{I}_{B'}}{|B|}.\end{aligned}$$

Optimization Program 2.

Here the last condition follows from the Choi condition Eq. (15) because we have eliminated the states $\hat{C}_{0,\hat{A}A'}$ and $\hat{C}_{0,\hat{B}B'}$ from explicit consideration.

(2) The other case is the one of the nonlocal success flags; e.g., Alice and Bob succeed if $f_A = f_B$. This is the case for example for the BBPSSW [8] or DEJMPS [9] protocols (again see also Appendix B 1). In this case we obtain

$$\begin{aligned}\text{maximize}\quad & \frac{|A||B|}{\delta}\,\mathrm{tr}\left[|\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}}\otimes\rho^T_{A'B'}\right.\\ & \left.\times(\hat{C}_{1,\hat{A}A'}\otimes\hat{C}_{1,\hat{B}B'}+\hat{C}_{0,\hat{A}A'}\otimes\hat{C}_{0,\hat{B}B'})\right]\\ \text{subject to}\quad & |A||B|\,\mathrm{tr}\left[\rho^T_{A'B'}(\hat{C}_{1,A'}\otimes\hat{C}_{1,B'}+\hat{C}_{0,A'}\otimes\hat{C}_{0,B'})\right]\\ & =\delta,\\ & \hat{C}_{1,\hat{A}A'}\geqslant 0,\quad \hat{C}_{1,\hat{B}B'}\geqslant 0,\quad \hat{C}_{0,\hat{A}A'}\geqslant 0,\\ & \hat{C}_{0,\hat{B}B'}\geqslant 0,\quad \hat{C}_{1,A'}+\hat{C}_{0,A'}=\frac{\mathbb{I}_{A'}}{|A|},\\ & \hat{C}_{1,B'}+\hat{C}_{0,A'}=\frac{\mathbb{I}_{B'}}{|B|}.\end{aligned}$$

Optimization Program 3.

C. Reliable upper bounds using SDP relaxations

The Choi isomorphism only transfers the optimization from channel space to state space, but it does not deal with the (quadratic) nonconvex nature of the program. In this section we perform a set of convex relaxations on the domain of optimization. First, in Sec. III C 1 we consider optimization over positive partial transpose (PPT) operations and in Sec. III C 2 we add an additional constraint related to the extendibility of separable states. We will call the resulting bounds reliable, since these numerical methods are guaranteed to produce an upper bound on our objective function. In contrast, later in Sec. III D we discuss a heuristic method which does not have this property.

1. PPT relaxations

The first method to obtain an upper bound on the objective is a direct extension of Rains [21]. Here, we relax the set of SEP-C states to the set of PPT Choi states—Choi states which are positive under partial transpose. We perform an easy adaption of this method to the case of MX operations including classical flags, resulting in Optimization Program 4. This method is implemented in our numerical software package available at [33].

Enforcing the PPT condition is an SDP constraint, whereas membership of SEP is more difficult to characterize and optimization over the set of separable states is in general hard. Applying the PPT constraint to our problem means that we construct a single Choi state variable on all the registers, such that it obeys the PPT condition, i.e.,

$$C^{\Gamma}_{\hat{A}F_AA'\hat{B}F_BB'}\geqslant 0, \tag{23}$$

where Γ denotes the transpose on all the registers of Bob.

To introduce some helpful notation, we can split this Choi of the distillation channel into the success and failure parts

$$C_{\hat{A}F_AA'\hat{B}F_BB'}=\hat{C}_{\checkmark,\hat{A}F_AA'\hat{B}F_BB'}+\hat{C}_{7,\hat{A}F_AA'\hat{B}F_BB'} \tag{24}$$

obeying the condition

$$\hat{C}_{\checkmark,A'B'}+\hat{C}_{\text{✗},A'B'}=\frac{\mathbb{I}_{A'B'}}{|A||B|}. \tag{25}$$

For a protocol with local flags we have

$$\hat{C}_{\checkmark,\hat{A}F_AA'\hat{B}F_BB'}=\hat{C}_{1,\hat{A}A'}\otimes\hat{C}_{1,\hat{B}B'}\otimes|11\rangle\langle 11|_{F_AF_B}, \tag{26}$$

whereas for a protocol with nonlocal flags

$$\begin{aligned}\hat{C}_{\checkmark,\hat{A}F_AA'\hat{B}F_BB'}&=\hat{C}_{1,\hat{A}A'}\otimes\hat{C}_{1,\hat{B}B'}\otimes|11\rangle\langle 11|_{F_AF_B}\\ &\quad+\hat{C}_{0,\hat{A}A'}\otimes\hat{C}_{0,\hat{B}B'}\otimes|00\rangle\langle 00|_{F_AF_B}.\end{aligned} \tag{27}$$

Clearly $\hat{C}_{\checkmark,\hat{A}F_AA'\hat{B}F_BB'}$ and $\hat{C}_{\text{✗},\hat{A}F_AA'\hat{B}F_BB'}$ are orthogonal on the flag registers. As a result imposing the PPT constraint on $C_{\hat{A}F_AA'\hat{B}F_BB'}$ is equivalent to imposing it on both $\hat{C}_{\checkmark,\hat{A}F_AA'\hat{B}F_BB'}$ and $\hat{C}_{\text{✗},\hat{A}F_AA'\hat{B}F_BB'}$. Finally, $\hat{C}_{\text{✗},\hat{A}F_AA'\hat{B}F_BB'}$ does not appear explicitly in our optimization problem, but because of the relation in Eq. (25), it translates directly to the following condition on the marginal of $\hat{C}_{\checkmark,\hat{A}F_AA'\hat{B}F_BB'}$:

$$\hat{C}^{\Gamma}_{\checkmark,A'B'}\leqslant\frac{\mathbb{I}_{A'B'}}{|A||B|}, \tag{28}$$

where Γ again denotes the partial transpose on all registers of B. Of course Eq. (25) also implies that

$$\hat{C}_{\checkmark,A'B'}\leqslant\frac{\mathbb{I}_{A'B'}}{|A||B|}. \tag{29}$$

Since in our program we have already eliminated the flags, our SDP variable is $\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}$. We note that both the cases with local and nonlocal flags as well as any other flag configuration reduce to exactly the same relaxed PPT program. All other constraints in terms of the reduced state of $\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}$ remain the same so that now we will obtain the following program:

$$\begin{aligned}\text{maximize}\quad & \frac{|A||B|}{\delta}\,\mathrm{tr}\left[\left(|\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}}\otimes\rho^T_{A'B'}\right)\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}\right]\\ \text{subject to}\quad & |A||B|\,\mathrm{tr}\left[\left(\mathbb{I}_{\hat{A}\hat{B}}\otimes\rho^T_{A'B'}\right)\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}\right]=\delta,\end{aligned}$$

$$
\begin{aligned}
&\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'} \geqslant 0,\\
&\hat{C}^{\Gamma}_{\checkmark,\hat{A}A'\hat{B}B'} \geqslant 0,\\
&\hat{C}_{\checkmark,A'B'} \leqslant \frac{\mathbb{I}_{A'B'}}{|A||B|},\\
&\hat{C}^{\Gamma}_{\checkmark,A'B'} \leqslant \frac{\mathbb{I}_{A'B'}}{|A||B|}.
\end{aligned}
$$

Optimization Program 4.

We give a side remark regarding terminologies. Such a PPT Choi state $C_{\hat{A}F_A A'\hat{B}F_B B'}$ corresponds to an operation that Rains defines as a PPT operation [21,35,36]. These PPT operations include all LOCC operations as a strict subset. Hence our relaxed program provides upper bounds on the achievable fidelity not only over MX and MXS operations but also over all LOCC operations. See Appendix A for a short discussion of these PPT channels.

The Optimization Program 4 is a semidefinite program with very high symmetry. This allows considerable further simplifications (see Appendix C). We finally obtain the semidefinite program corresponding to the Rains style bound on the fidelity of distillation with fixed success probability δ:

$$
\begin{aligned}
\text{maximize}\quad & p(M_{A'B'},E_{A'B'}) = \frac{|A||B|}{\delta}\,\mathrm{tr}\left[\rho^{T}_{A'B'}M_{A'B'}\right]\\
\text{subject to}\quad & M_{A'B'} \geqslant 0,\quad E_{A'B'} \geqslant 0,\\
& M_{A'B'} + E_{A'B'} \leqslant \frac{\mathbb{I}_{A'B'}}{|A||B|},\\
& M^{\Gamma}_{A'B'} + E^{\Gamma}_{A'B'} \leqslant \frac{\mathbb{I}_{A'B'}}{|A||B|},\\
& |A||B|\,\mathrm{tr}\left[\rho^{T}_{A'B'}(M_{A'B'}+E_{A'B'})\right] = \delta,\\
& M^{\Gamma}_{A'B'} + \frac{1}{D+1}E^{\Gamma}_{A'B'} \geqslant 0,\\
& -M^{\Gamma}_{A'B'} + \frac{1}{D-1}E^{\Gamma}_{A'B'} \geqslant 0.
\end{aligned}
$$

Optimization Program 5.

Recall that $\rho_{A'B'}$ is the initial input state that Alice and Bob are attempting to distill and in most examples considered here, it will consist of two copies of some two-qubit state. In what follows and on all the plots shown in Sec. IV we will refer to the bound obtained using this program as the *PPT bound*.

We note here that by following an analogous procedure, one can construct a similar program which aims at maximizing probability of success subject to a constraint of fixed output fidelity. This program can also be relaxed to a PPT program which is also an SDP. Effectively it results in a similar program to the one above just with the objective function and constraint on probability of success interchanged:

$$
\begin{aligned}
\text{maximize}\quad & |A||B|\,\mathrm{tr}\left[\rho^{T}_{A'B'}(M_{A'B'}+E_{A'B'})\right]\\
\text{subject to}\quad & M_{A'B'} \geqslant 0,\quad E_{A'B'} \geqslant 0,\\
& M_{A'B'} + E_{A'B'} \leqslant \frac{\mathbb{I}_{A'B'}}{|A||B|},\\
& M^{\Gamma}_{A'B'} + E^{\Gamma}_{A'B'} \leqslant \frac{\mathbb{I}_{A'B'}}{|A||B|},\\
& \mathrm{tr}\left\{\rho^{T}_{A'B'}[(1-F)M_{A'B'} - FE_{A'B'}]\right\} = 0,\\
& M^{\Gamma}_{A'B'} + \frac{1}{D+1}E^{\Gamma}_{A'B'} \geqslant 0,\\
& -M^{\Gamma}_{A'B'} + \frac{1}{D-1}E^{\Gamma}_{A'B'} \geqslant 0.
\end{aligned}
$$

Optimization Program 6.

Now F is a constant fidelity and so the fidelity constraint is just

$$
\frac{\mathrm{tr}\left[\rho^{T}_{A'B'}M_{A'B'}\right]}{\mathrm{tr}\left[\rho^{T}_{A'B'}(M_{A'B'}+E_{A'B'})\right]} = F. \tag{30}
$$

Hereafter, we will drop the subscripts on ρ, E, and M to simplify the notation.

We remark that an appealing feature of semidefinite programs is the dual [37] of the SDP. In Appendix D we dualize the above SDPs to obtain dual programs which depend on the variables y,J,G,H,K. We denote the objective function of the dual program as $d(y,J,G,H,K)$. It is an appealing feature of SDP duality—known as *weak duality*—that

$$
d(y,J,G,H,K) - p(M,E) \geqslant 0. \tag{31}
$$

Finding values for y,J,G,H, and K that satisfy the constraints of the dual SDP thus always results in upper bounds $d(y,J,G,H,K) \geqslant p^{*}$, where p^{*} denotes the optimal solution of the primal program. Furthermore, if such variables satisfy $d(y,J,G,H,K) = p(M,E)$, then we know that the optimal solution has been found.

We remark that it is this feature that makes SDPs highly appealing as a numerical method, since a numerical SDP solver will find primal and dual variables which form a certificate for optimality, or—if due to finite precision in numerical calculations optimality is reached only approximately—a certificate for approximate optimality in which the difference between the dual and primal $(d-p)$ is sufficiently small. In addition, however, SDPs can thus also be used to prove optimality analytically, if one can make an educated guess for the primal and dual variables.

2. Bose symmetric extensions

The goodness of the relaxation above depends on how well the set of PPT Choi states approximates the set SEP-C. A sharper approximation could evidently be obtained by approximating the set of separable states SEP itself by more stringent conditions. A standard technique for doing so is by the method of extensions [31,32] which is closely related to the sums-of-squares relaxations for polynomial optimization problems.

In the case at hand, in addition to the PPT constraint in Eq. (23) we will add the constraint that the state is k-Bose-symmetric-extendible (k-BSE) [38]. By definition, a (Choi) state $\hat{C}_{(\hat{A}A')\hat{B}B'}$ is k-BSE iff there exists $\hat{C}_{(\hat{A}_1A'_1)\ldots(\hat{A}_{k+1}A'_{k+1})\hat{B}B'}$ satisfying

(1) $\hat{C}_{(\hat{A}_1A'_1)\ldots(\hat{A}_{k+1}A'_{k+1})\hat{B}B'} \geqslant 0$,

(2) $\mathrm{tr}_{(\hat{A}_2A'_2)\ldots(\hat{A}_{k+1}A'_{k+1})}(\hat{C}_{(\hat{A}_1A'_1)\ldots(\hat{A}_{k+1}A'_{k+1})\hat{B}B'}) = \hat{C}_{(\hat{A}A')\hat{B}B'}$,

(3) $(P_{\mathrm{Sym}} \otimes \mathbb{I}_{\hat{B}B'})(\hat{C}_{(\hat{A}_1A'_1)\ldots(\hat{A}_{k+1}A'_{k+1})\hat{B}B'}) =$ $\hat{C}_{(\hat{A}_1A'_1)\ldots(\hat{A}_{k+1}A'_{k+1})\hat{B}B'}$, where P_{Sym} is the projector onto the symmetric subspace of $(\hat{A}_1A'_1)\ldots(\hat{A}_{k+1}A'_{k+1})$.

It is clear that adding this constraint to the PPT constraint constitutes a sharper approximation of SEP-C because any separable state is k-BSE for all $k \in \mathbb{N}$. To see this, it is sufficient to note that $\sum_i p_i |u_i\rangle\langle u_i|^{\otimes k+1} \otimes |v_i\rangle\langle v_i|$ is a k Bose symmetric extension of the separable state $\sum_i p_i |u_i\rangle\langle u_i| \otimes |v_i\rangle\langle v_i|$.

In this way, we obtain a sharper and sharper approximation of SEP-C by choosing larger values of k—the accuracy scales not worse than $O(|\hat{A}A'|^2/(k+1)^2)$ [39]. The only drawback is the size of the resulting SDP. Although it increases only polynomially with k, for practically interesting problems we were only able to introduce $k = 1$ Bose symmetric extensions. We refer to Appendix E for the detailed calculations and the exact form of the resulting SDP. Whenever we refer to the *1-BSE bound*, we mean the bound arising from this optimization over Choi matrices that are both PPT and 1-BSE.

D. Optimizing existing schemes

While the previous methods are concerned with deriving upper bounds on the fidelity, we can as well start from an existing distillation protocol and try to find a better protocol. In the following we discuss one such a scheme that we dub the seesaw method. Looking at the original Optimization Programs 2 and 3, we see that there is no need for any PPT style relaxation if one of the distillation maps for either Alice or Bob is fixed: for a fixed value of one of the maps, the optimization problem is already an SDP. If we thus fix the operation of Alice (or Bob), then we may use an SDP solver to optimize over the possible distillation schemes in terms of the Choi state of Bob (or Alice). Once solved, we may iterate the procedure in a seesaw fashion. We now fix the operation of Bob (Alice) with the outcome of the previous step and we optimize over the operation of Alice (Bob). The optimization problem is again an SDP. These steps can then be repeated, as often as desired optimizing iteratively over either Alice or Bob. While not guaranteed to find the optimal solution, the seesaw method often performs rather well in practice and is implemented in our numerical package [33]. In fact, in the next section we provide an example where this method finds an optimal filtering scheme, as the numerical results show that it achieves fidelities corresponding to the PPT bound. We remark that given the new Choi states, one may find the corresponding isometry (or unitary) that implements the map using an ancilla (see, e.g., lecture notes [40]) and then compile it into a quantum circuit for the specific architecture in question.

IV. STATES AND DISTILLATION SCHEMES

Let us now illustrate our methods with a number of states commonly studied in the entanglement distillation literature, or arising in experiments. We thereby demonstrate the use of our methods as a numerical tool to compute the trade-offs between the fidelity F and probability of success p_{succ}, as well as their use as an analytical tool to formally prove optimality of certain entanglement distillation schemes. We also provide a simple example illustrating the use of the seesaw method to improve an existing distillation scheme for a specific state.

Here we will use the term "a copy of a state" to denote a two-qubit state shared between Alice and Bob. In these examples, we will for simplicity only consider distillation to a single copy, i.e., when the output of the procedure is a two-qubit state. More examples can easily be explored using the freely available numerical package [33].

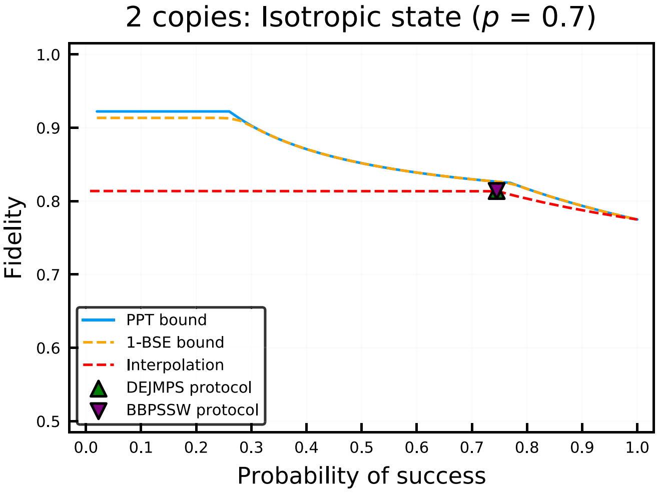

FIG. 1. Distilling the isotropic states $\tau_{ab}^{\otimes 2}$ with $D = 2$ and $p = 0.7$ in Eq. (32) to a two-qubit state. The fidelity of each input copy is $F_{\text{in}} = 0.775$ and we observe that deterministic distillation (with $p_{\text{succ}} = 1$) is not possible for two copies of the isotropic state. We also find that the method of 1-BSE provides tighter bounds than the PPT method alone.

A. Isotropic states

As a warm-up, let us consider distilling isotropic states. These states are often considered in the quantum information theory literature due to their beautiful symmetries. Moreover, they are the states that arise when a maximally entangled state undergoes depolarizing noise, which is often used as a simplified pessimistic model for the noise caused by the imperfect operations in physical implementations of quantum memories. Specifically, an isotropic state is of the form

$$\tau_{AB} = p|\Phi_D\rangle\langle\Phi_D| + (1-p)\frac{\mathbb{I}}{D^2}, \qquad (32)$$

where $|\Phi_D\rangle$ is the maximally entangled state defined in Eq. (2). The isotropic state is invariant under $U \otimes U^*$ on A and B for all U.

Numerical examples

Figure 1 illustrates the upper bounds obtained by PPT and the 1-BSE relaxation, in comparison to the BBPSSW and DEJMPS protocols when distilling 2 copies of the isotropic state $\rho_{AB} = \tau_{ab}^{\otimes 2}$ to a single two-qubit state (see Appendix B 1 for the description of these well-known protocols). We remark that when performing a single round of distillation, the two protocols coincide for the case of the isotropic state. The continuous red line corresponds to an achievable scheme based on the interpolation or extrapolation of those existing schemes. The details of how this is performed are included in Appendix B 2 and for simplicity on the plots we always label this curve arising from both extrapolation and interpolation as "Interpolation". Similarly in Fig. 2 we depict the corresponding

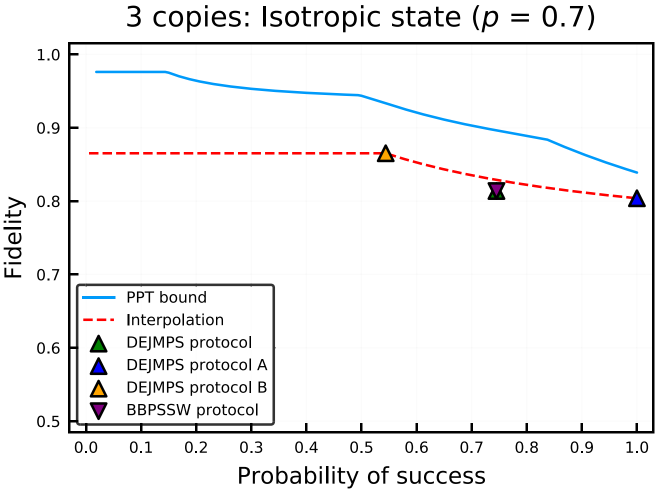

FIG. 2. Distilling the isotropic states $\tau_{ab}^{\otimes 3}$ with $D = 2$ and $p = 0.7$ in Eq. (32) to a two-qubit state. The fidelity of each input copy is $F_{\text{in}} = 0.775$. The protocol DEJMPS A corresponds to applying DEJMPS to the first two copies and outputting the resulting state in case of success and outputting the remaining third copy in case of failure. This protocol allows for deterministic increase of fidelity. The protocol DEJMPS B corresponds to applying DEJMPS to the first two copies and then conditioned on success, applying it to the remaining two copies. Failure at any stage results in outputting the failure flag. The 1-BSE bound was already computationally too expensive for this 3-copy scenario.

results for distilling 3 copies of the isotropic state $\rho_{AB} = \tau_{ab}^{\otimes 3}$ to a two-qubit state.

In Figs. 1 and 2 we see that both the PPT and 1-BSE bounds are nontrivial and the 1-BSE bound is tighter than the PPT bound for smaller values of the probability of success. In particular we observe that deterministic distillation (with $p_{\text{succ}} = 1$) when operating on 2 copies of the isotropic state is not possible. For 3 copies it is possible to deterministically increase the fidelity, and this can be achieved, e.g., using the protocol DEJMPS A (see caption of Fig. 2 for details of this protocol).

B. Bell diagonal states

More generally, we now consider states τ_{AB} that are diagonal in the Bell basis given by

$$|\Phi^+\rangle = |\Phi_2\rangle, \tag{33}$$

$$|\Phi^-\rangle = (\mathbb{I} \otimes Z)|\Phi_2\rangle, \tag{34}$$

$$|\Psi^+\rangle = (\mathbb{I} \otimes X)|\Phi_2\rangle, \tag{35}$$

$$|\Psi^-\rangle = (\mathbb{I} \otimes XZ)|\Phi_2\rangle. \tag{36}$$

These are interesting states to consider since indeed any two-qubit state ρ_{AB} can be brought into this form by twirling it over the group of correlated Pauli operators: $\{X \otimes X, Y \otimes Y, Z \otimes Z, \mathbb{I} \otimes \mathbb{I}\}$. This can be achieved if Alice and Bob have access to some shared randomness. We can thus consider entangled states

$$\begin{aligned}\tau_{AB} = {} & p_1|\Phi^+\rangle\langle\Phi^+| + p_2|\Psi^+\rangle\langle\Psi^+| + p_3|\Phi^-\rangle\langle\Phi^-| \\ & + (1 - p_1 - p_2 - p_3)|\Psi^-\rangle\langle\Psi^-|,\end{aligned} \tag{37}$$

where $p_1 > 0.5$ and $p_1 > p_2 \geqslant p_3 \geqslant 1 - p_1 - p_2 - p_3$. Any Bell diagonal state for which one of the Bell coefficients is larger than 0.5 can be rotated into this form using only local Clifford operations performed by Alice and Bob.

The distillation of such states has been studied in the literature, and we will focus here on the action of the DEJMPS protocol on these states since it is known for achieving higher fidelities than the BBPSSW protocol. Specifically, Alice and Bob share two copies of a Bell diagonal state τ_{AB}, that is, $\rho_{AB} = \tau_{ab}^{\otimes 2}$. The decreasing order of the Bell coefficients in τ_{AB} is important as this specific ordering allows us to achieve the highest fidelity over all the orderings [41].

We note that it has been recently shown that the DEJMPS protocol achieves the highest possible fidelity over LOCC operations when distilling a two-qubit state from two copies of a Bell diagonal state of rank two [42]. Moreover, in Ref. [41] protocols that permute Bell states in the mixture were analyzed and it was claimed that for two copies of all Bell diagonal states, DEJMPS protocol achieves the highest achievable fidelity when distilling a two-qubit state, but only among all such permuting protocols. Here our results indicate that we can make a much wider range of optimality statements about DEJMPS in relation to Bell diagonal states than has been known before.

1. Numerical examples

We first investigate a number of examples using our numerical procedure. We present the results in Fig. 3 and in Fig. 4. We again emphasize that for simplicity we only consider distilling a two-qubit state from two copies of a Bell diagonal state and we note that all these optimality statements apply when optimizing over all LOCC protocols.

First, we observe that for all Bell diagonal states of rank up to three DEJMPS achieves the highest possible output fidelity and achieves it with the highest possible probability of success, as can be seen in a specific example in Fig. 3. This statement we also prove analytically as described in the next subsection. Moreover, as we also illustrate in Fig. 3, we numerically observe that for Bell diagonal states of rank up to three, extrapolating from DEJMPS allows us to achieve the highest possible output fidelity for each extrapolation protocol's probability of success.

Finally, we also numerically observe that for Bell diagonal states of rank four, apart from a certain set of states including and around the isotropic state, DEJMPS achieves the highest possible fidelity for this protocol's probability of success when applied to these states. In Fig. 5 we fix p_1 and p_2 and plot the gap between our numerical upper bound and the output fidelity of DEJMPS, both evaluated at the probability of success of DEJMPS, versus the parameter p_3. We see that in this space of Bell coefficients the gap vanishes when one moves far enough from the isotropic state. In this space, we observe a similar gap in any other direction away from the isotropic state. However, only by moving exactly along the axis of one of those coefficients do we obtain a gap that is symmetric around the

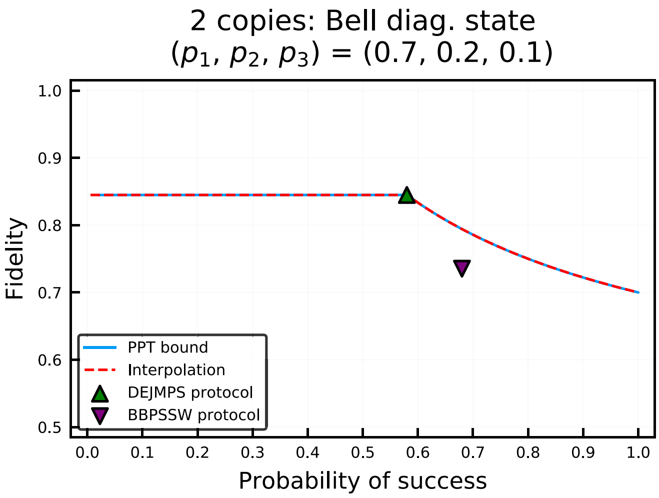

FIG. 3. Distilling the Bell diagonal states of rank-three $\tau_{ab}^{\otimes 2}$ with $D = 2$ and $p_1 = 0.7, p_2 = 0.2, p_3 = 0.1$ in Eq. (37) to a two-qubit state. The fidelity of each input copy is $F_{\text{in}} = 0.7$ and we observe that deterministic distillation (with $p_{\text{succ}} = 1$) is not possible for two copies of this state. We see that DEJMPS is optimal for a mixture of three Bell states. Moreover, extrapolating from DEJMPS to higher probability of success as described in Appendix B 2, we see that the extrapolation curve overlaps with the PPT bound for all values of the probability of success. This means that this extrapolation also results in optimal schemes achieving the highest possible output fidelity for the specific fixed probability of success. The 1-BSE bound is not included because it overlaps with the PPT bound.

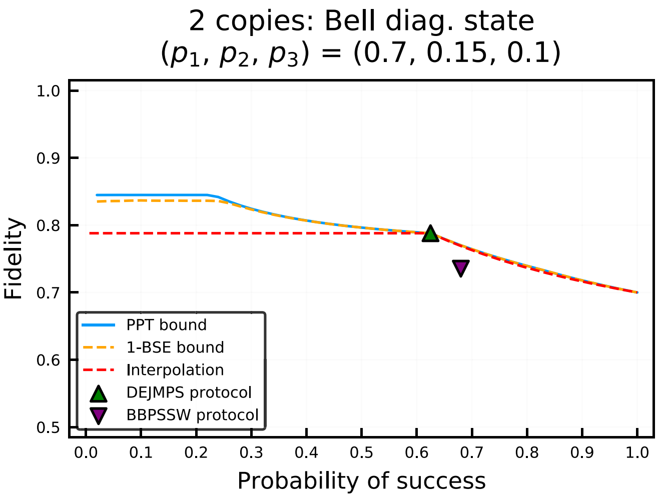

FIG. 4. Distilling the Bell diagonal states of rank-four $\tau_{ab}^{\otimes 2}$ with $D = 2$ and $p_1 = 0.7, p_2 = 0.15, p_3 = 0.1$ in Eq. (37) to a two-qubit state. The fidelity of each input copy is $F_{\text{in}} = 0.7$ and we observe that deterministic distillation (with $p_{\text{succ}} = 1$) is not possible for two copies of this state. We also find that the 1-BSE bound is tighter than the PPT bound for smaller values of the probability of success. Finally, we observe that DEJMPS achieves the highest possible output fidelity for this protocol's probability of success for a mixture of four Bell states which are far enough from the isotropic state.

isotropic state as in Fig. 5. The reason for this fact is that on those axes the two states that are located symmetrically on two sides of the peak at the isotropic state are the same up to the permutation of the Bell coefficients.

2. *Optimal fidelity and success probability*

Semidefinite programming duality now allows us to prove analytically that DEJMPS is an optimal protocol for distilling from two copies of all Bell diagonal states of rank up to three, which was not known before.

Theorem 1 (informal). Given two copies of a Bell diagonal state of rank at most three and distillation towards the target maximally entangled state with $D = 2$, there is no protocol that achieves a larger fidelity than DEJMPS and there is no protocol that achieves this fidelity with a larger success probability than DEJMPS.

In the following we sketch the proof of Theorem 1. We leave the full details including a precise definition of optimality to Appendix G.

The entangled Bell diagonal states of rank up to three can be written as

$$\begin{aligned}\tau_{AB} = {} & p_1|\Phi^+\rangle\langle\Phi^+| + p_2|\Psi^+\rangle\langle\Psi^+| \\ & + (1 - p_1 - p_2)|\Phi^-\rangle\langle\Phi^-|, \qquad (38)\end{aligned}$$

with $p_1 > 0.5$ and $p_1 > p_2 \geqslant 1 - p_1 - p_2$. First we note that the DEJMPS protocol applied to two copies of the state in

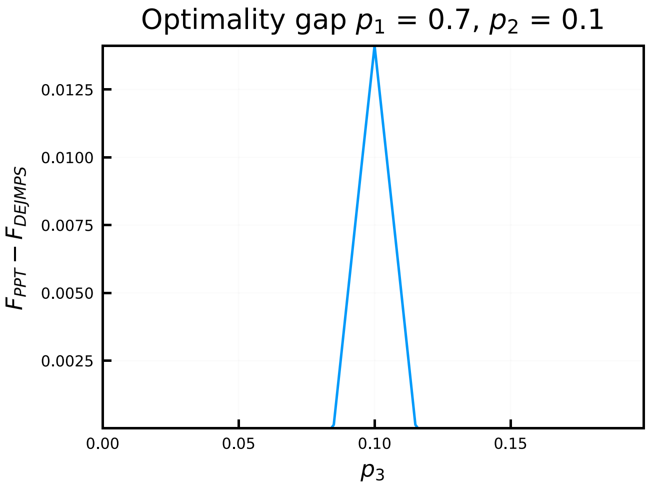

FIG. 5. Distilling the Bell diagonal states of rank-four $\tau_{ab}^{\otimes 2}$ with $D = 2$ and $p_1 = 0.7, p_2 = 0.1$ in Eq. (37) to a two-qubit state. The fidelity of each input copy is $F_{\text{in}} = 0.7$. The plot shows the difference between the PPT bound and the fidelity achievable through DEJMPS as a function of p_3 for the probability of success of DEJMPS. We see that DEJMPS achieves the highest possible output fidelity for this protocol's probability of success for a mixture of four Bell states which are far enough from the isotropic state (the middle of the peak). Clearly the states considered on this plot for which $p_3 \neq 0.1$ do not satisfy the condition $p_1 > p_2 \geqslant p_3 \geqslant 1 - p_1 - p_2 - p_3$; therefore when applying the DEJMPS protocol to such a state we first permute the Bell coefficients to this order. The 1-BSE bound is not included because it overlaps with the PPT bound.

Eq. (38) conditioned on success results in a state

$$\rho_{\hat{A}\hat{B}} = p_1'|\Phi^+\rangle\langle\Phi^+| + p_2'|\Psi^+\rangle\langle\Psi^+| + (1 - p_1' - p_2')|\Psi^-\rangle\langle\Psi^-|, \tag{39}$$

where

$$p_1' = \frac{p_1^2}{N}, \tag{40}$$

$$p_2' = \frac{p_2^2 + (1 - p_1 - p_2)^2}{N}, \tag{41}$$

and $N = p_1^2 + (1 - p_1)^2$ is the probability that the protocol succeeds. Note that $p_1' > p_2' \geqslant 1 - p_1' - p_2'$. Moreover the fidelity increases, that is, $p_1' > p_1$.

The strategy to show optimal fidelity relies on the dual formulation of the SDP in Optimization Program 5. In particular, we prove that there exists a feasible solution of the dual program with the objective function value corresponding to p_1' for all $\delta \in (0,1]$. Hence p_1' is an upper bound on the achievable fidelity for all δ and there cannot exist an LOCC protocol that takes two copies of the state in Eq. (38) and outputs a state with fidelity larger than p_1'.

The proof of N being the optimal success probability for all protocols that output fidelity equal to p_1' also follows from SDP duality. That is, we show that there exists a feasible solution of the dual program for optimizing the probability of success with the objective function taking the value N for the output fidelity $F = p_1'$.

C. R states

Another interesting class of states are quantum states that form a mixture between a maximally entangled state and a product state. In particular let us first consider a case where the product part of the mixture is orthogonal to the maximally entangled part. Specifically let us consider the state

$$\tau_{AB} = p|\Psi^\pm\rangle\langle\Psi^\pm| + (1 - p)|11\rangle\langle 11|, \tag{42}$$

which we will call an R state. We note that up to a local X or XZ gate this state is exactly the state in Eq. (10) that we considered in the filtering example in Sec. III A (this local flip on one side will be helpful when discussing remote entanglement generation in the following section).

This type of state is interesting for two reasons. The first one is "mathematical." The above R state is a simple example of a state that as expressed in Ref. [43] possesses local information, in the sense that the reduced state on Alice and Bob individually is not a maximally mixed state. This local information can also be seen in the nonzero off-diagonal elements when the state is expressed in the Bell basis. Since for the DEJMPS and BBPSSW protocols the output fidelity and probability of success are completely independent of those off-diagonal coefficients, this local information is completely neglected in those protocols. Hence one could expect that for these states there exist distillation strategies that utilize this local information and in this way possibly outperform the DEJMPS protocol.

As observed in Ref. [20] this is indeed the case, since for any value of $0 < p \leqslant 1$ it is possible to extract a perfect Bell state from two copies of the R state by performing a bilateral CNOT, measuring the target copy in the standard basis and postselecting the events for which both Alice and Bob measured the target copy to be one. In such a scenario of applying this protocol to two copies of the R state the fidelity of $F = 1$ is achieved with probability of success $p_{\text{succ}} = p^2/2$. Note that depending on the value of p the R state might actually have fidelity to any maximally entangled state smaller than or equal to half. This shows a fundamental difference with respect to the protocols that do not utilize this local information like DEJMPS or BBPSSW for which it is required that the initial fidelity to some maximally entangled state is larger than 0.5 [57].

The second reason for considering these states is experimental. These states arise in certain protocols for remote entanglement generation that use a single photon detection scheme in the presence of photon loss [7,13,44]. In particular, [7] describes an entanglement generation procedure that generates two copies of a state closely related to the R state (see the next section for more details) and then performs the above described distillation protocol proposed in Ref. [20] to combat the effect of photon loss. Since the authors refer to this entire entanglement generation scheme as the Extreme Photon Loss scheme (EPL), here we will refer to this distillation protocol used within the EPL procedure as EPL-D. As already mentioned and as we will discuss in the next section, the R state is still only an idealization of the actual raw state generated within the remote entanglement generation schemes described in Refs. [7,13]. In particular the R state includes only noise due to the photon loss while all realistic implementations will also suffer from other types of noise.

Numerical examples

We first look at filtering a single copy of the R state, since as stated in Sec. III A, there exists a well-known protocol for filtering those states. Optimal filtering schemes have been studied in the literature [45–47], but not in the context of the optimal trade-off of fidelity and probability of success.

First, we note that the filtering scheme described in Sec. III A [here we assume that before filtering, Alice applies an X or XZ operation to bring the R state to the form in Eq. (10)] clearly cannot increase the fidelity deterministically, while from [47] we know that for all $p < 2/3$ there exists a way of deterministically increasing the fidelity of the R state by running a probabilistic filtering protocol and outputting a product state of fidelity half in case of failure. Inspired by this result we consider here a modified version of the discussed filtering scheme in which for certain larger values of the desired success probability for R states with $p < 2/3$, conditioned on the failure of that original scheme Alice and Bob probabilistically output a state of fidelity half. The details of this modification are discussed in Appendix B 2. In Fig. 6 and in Fig. 7 we compare this modified filtering scheme with our numerical bounds. We consider one example for which the input fidelity is larger and one for which it is smaller than half.

The original filtering scheme allows us to choose the desired probability of success by making a suitable choice of the ϵ parameter, while in our modified scheme success probability can also be varied by changing the probability of outputting a product state in case of failure of the original scheme (here we maximize the fidelity over those two parameters for each

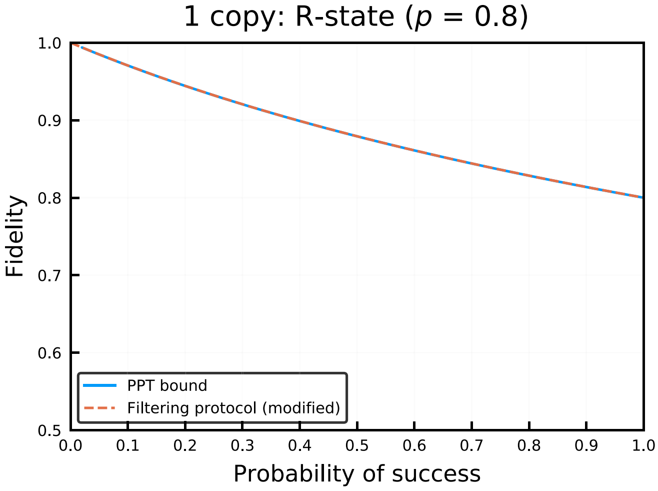

FIG. 6. Filtering R state τ_{AB} with $D = 2$ and $p = 0.8$ in Eq. (42) to a two-qubit state. The fidelity of the input copy is $F_{\text{in}} = 0.8$ and in accordance with [47] we observe that deterministic filtering (with $p_{\text{succ}} = 1$) is not possible for this state. We see that the filtering scheme perfectly overlaps with the PPT bound, which proves its optimality for this state. The 1-BSE bound is not included because it overlaps with the PPT bound.

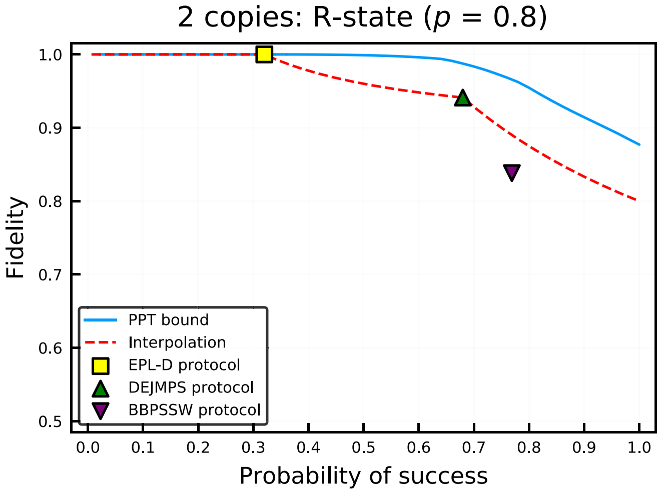

FIG. 8. Distilling the R states $\tau_{ab}^{\otimes 2}$ with $D = 2$ and $p = 0.8$ in Eq. (42) to a two-qubit state. The fidelity of the input copy is $F_{\text{in}} = 0.8$ and we observe that while the extrapolation from DEJMPS does not allow for deterministic distillation (with $p_{\text{succ}} = 1$) in this case, the PPT bound still allows for this possibility. We also see that EPL-D allows for achieving unit fidelity. The 1-BSE bound is not included because it overlaps with the PPT bound.

probability of success). We note that independently of the value of the parameter p (provided that it is nonzero), in the limit of zero success probability, this filtering scheme allows for obtaining a state that is arbitrarily close to a maximally entangled state. From the numerical results we observe that for the considered values of p, we have that for all probabilities of success our PPT bound perfectly overlaps with the modified filtering scheme, proving that no higher fidelity can be achieved for the fixed value of probability of success than already achieved by our modified filtering scheme. Hence the modified filtering scheme is in fact optimal for these states.

We also present two numerical examples for distillation from two to one copies of the R state in Fig. 8 and in Fig. 9. In Fig. 8 we consider two copies of the R state with input fidelity of 0.8. We see that while our achievable interpolation scheme cannot deterministically increase fidelity for this state, the nontrivial numerical bounds still allow for this possibility. We also see that for this state the PPT operations allow for distilling

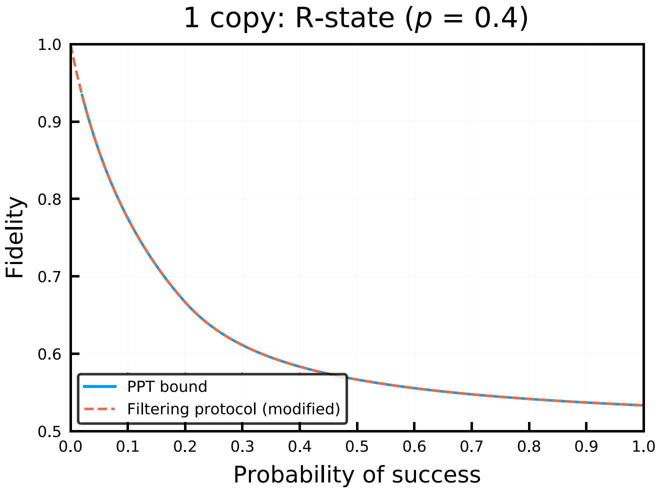

FIG. 7. Filtering R state τ_{AB} with $D = 2$ and $p = 0.4$ in Eq. (42) to a two-qubit state. The fidelity of the input copy is $F_{\text{in}} = 0.4$. As first shown in Ref. [47], we observe that for the smaller values of p deterministic filtering of R states is possible and can be achieved with our scheme. We also see that the filtering scheme perfectly overlaps with the PPT bound, which proves its optimality for this state. The 1-BSE bound is not included because it overlaps with the PPT bound.

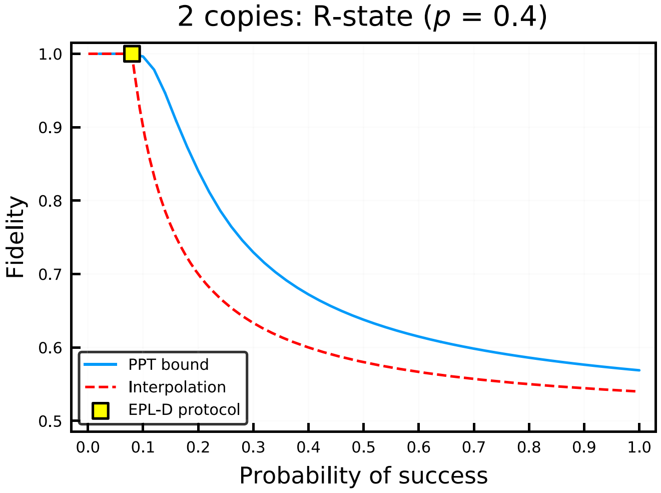

FIG. 9. Distilling the R states $\tau_{ab}^{\otimes 2}$ with $D = 2$ and $p = 0.4$ in Eq. (42) to a two-qubit state. The fidelity of the input copy is $F_{\text{in}} = 0.4$ and we observe that deterministic distillation (with $p_{\text{succ}} = 1$) which achieves output fidelity larger than half is easily achievable for two copies of this state. We also see that EPL-D allows for achieving unit fidelity even if $p \leqslant 0.5$. The 1-BSE bound is not included because it overlaps with the PPT bound.

a state very close to a maximally entangled state for much larger probability of success than the achievable interpolation scheme. In Fig. 9 we consider two copies of the R state whose input fidelity is smaller than half. In this case the interpolation scheme allows for deterministic increase of fidelity above 0.5 (as discussed in the previous paragraph, for this value of p that is possible even with just the modified filtering, but the interpolation scheme performs better). We see that here the PPT operations do not allow for distilling a state with fidelity close to one for probabilities of success much larger than that of the EPL-D protocol.

D. Remote entanglement generation

Here we expand on the experimentally relevant ideas described in the previous section on R states to reliably model the remote entanglement generation through distillation, including most of the experimentally relevant sources of noise as described in Ref. [7] and as realized experimentally in Ref. [18]. Specifically, in most experimental implementations of this specific entanglement generation scheme the actual state that is created will be of the form

$$\rho_{AB}(p) = \frac{1}{2\pi}\int d\phi \tau_{A1B1}(\phi,p)\otimes\tau_{A2B2}(\phi,p), \tag{43}$$

where

$$\tau_{AB}(\phi,p) = p|\Psi^+(\phi)\rangle\langle\Psi^+(\phi)| + (1-p)|11\rangle\langle 11|, \tag{44}$$

and

$$|\Psi^+(\phi)\rangle = \frac{1}{\sqrt{2}}(|01\rangle + e^{i\phi}|10\rangle), \tag{45}$$

$$|\Psi^-(\phi)\rangle = \frac{1}{\sqrt{2}}(|01\rangle - e^{i\phi}|10\rangle). \tag{46}$$

Here ϕ is a phase that arises due to the optical apparatus and in most cases is completely unknown. We see that the complete lack of knowledge of the phase ϕ leads to the uniform averaging over that phase. However, if the system is stable over the duration of generation of the two copies of ρ, one can assume that both of those copies are correlated in that phase.

In the next step we make this model even more precise by acknowledging the fact that the first copy of ρ will actually undergo dephasing while trying to generate the second copy. Moreover, the phase will in general not be exactly the same for both copies since in any realistic setting it could drift with respect to the first copy. Mathematically, those two effects can be combined together into a single dephasing process that affects one of the two copies

$$\rho_{AB}(p,p_d) = \frac{1}{2\pi}\int d\phi \tau_{A1B1}(\phi,p,p_d)\otimes\tau_{A2B2}(\phi,p,1), \tag{47}$$

where

$$\begin{aligned}\tau_{AB}(\phi,p,p_d) = p[&p_d|\Psi^+(\phi)\rangle\langle\Psi^+(\phi)| \\ &+(1-p_d)|\Psi^-(\phi)\rangle\langle\Psi^-(\phi)|] \\ &+(1-p)|11\rangle\langle 11|.\end{aligned} \tag{48}$$

Here we shall refer to the state in Eq. (47) as the "R-state correlated phase." In this scenario the successful implementation of the EPL-D distillation protocol (followed by a local rotation) leads to the output state

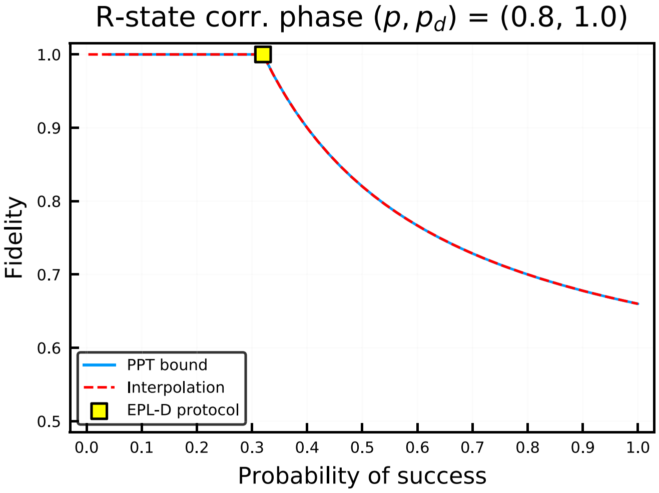

FIG. 10. Distilling the R-state correlated phase $\rho_{AB}(p,p_d)$ given in Eq. (47) with $D = 2$ and $p = 0.8, p_d = 1$ to a two-qubit state. We see that EPL-D is an optimal distillation protocol for the EPL remote entanglement generation scheme. The red extrapolation curve perfectly overlaps with the PPT bounds which means that the protocols arising by extrapolating EPL-D to higher values of probability of success are also optimal and achieve the maximum possible fidelity for the corresponding probability of success. The 1-BSE bound is not included because it overlaps with the PPT bound.

$$\eta_{\hat{A}\hat{B}}(p_d) = p_d|\Phi^+\rangle\langle\Phi^+| + (1-p_d)|\Phi^-\rangle\langle\Phi^-|, \tag{49}$$

with probability of success $p_{\text{succ}} = p^2/2$. We also provide a more detailed description of this remote entanglement generation scheme in Appendix B 1.

1. *Numerical examples*

We present two numerical examples for applying distillation to the state $\rho_{AB}(p,p_d)$ in Fig. 10 and in Fig. 11. We observe that EPL-D saturates the bound by achieving the highest possible fidelity with the highest possible probability of success. Moreover, we observe that extrapolating from EPL-D to higher values of probability of success also achieves the highest possible fidelity for the corresponding value of the probability of success.

2. *Optimal fidelity and probability of success*

The numerical examples suggest that the EPL-D protocol is optimal for distilling states $\rho_{AB}(p,p_d)$ given in Eq. (47) both in terms of output fidelity and probability of success. This means that the EPL scheme utilizes the optimal distillation protocol in this respect.

Theorem 2. Given a state of the form $\rho_{AB}(p,p_d)$ given in Eq. (47) and distillation towards the target maximally entangled state with $D = 2$, there is no protocol that achieves a larger fidelity than EPL-D and there is no protocol that achieves this fidelity with a larger success probability than EPL-D.

It turns out that in this case it is possible to analytically prove this optimality in a simple way without using the SDP formulation. Specifically, see Appendix H for the proof, that

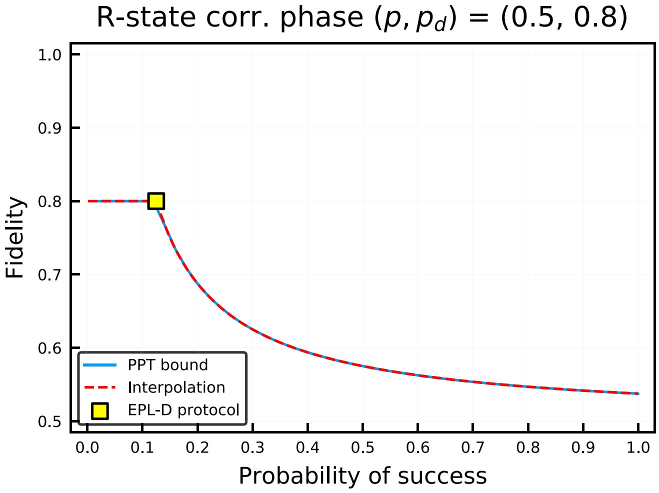

FIG. 11. Distilling the R-state correlated phase $\rho_{AB}(p,p_d)$ given in Eq. (47) with $D=2$ and $p=0.5, p_d=0.8$ to a two-qubit state. EPL-D is an optimal distillation protocol for the EPL remote entanglement generation scheme. The red extrapolation curve perfectly overlaps with the PPT bounds which means that the protocols arising by extrapolating EPL-D to higher values of probability of success are also optimal and achieve the maximum possible fidelity for the corresponding probability of success. The 1-BSE bound is not included because it overlaps with the PPT bound.

after performing the integration over the phase ϕ, the state $\rho_{AB}(p,p_d)$ is actually block diagonal in the standard basis, where one of the blocks is of size two and all the other blocks are of size one. Clearly the blocks of size one correspond to separable states. Hence, output fidelity is maximized by projecting onto the size-two block. Finally, this block is equivalent up to a local relabeling to the state $\eta_{\hat{A}\hat{B}}(p_d)$ in Eq. (49). Since this state is nonfilterable in the sense that even probabilistically no LOCC scheme can increase its fidelity [47], the optimal protocol cannot achieve fidelity higher than p_d which is achieved by EPL-D within the EPL scheme.

The same argument also implies that within EPL, EPL-D achieves fidelity p_d with maximum probability. More concretely, the probability of the projection onto the size-two block succeeds with probability at most $p^2/2$ which is the success probability of EPL-D within EPL.

3. *Optimality with respect to distillable entanglement*

Recall that the distillable entanglement of a state is defined as the optimal asymptotic rate at which it is possible to transform copies of the state into copies of the maximally entangled state. It turns out that within EPL, EPL-D is also optimal for distillable entanglement. More concretely:

Theorem 3. Given a state of the form $\rho_{AB}(p,p_d)$ given in Eq. (47), there is no protocol with the success probability of EPL-D that outputs a state with larger distillable entanglement. Equally there is no protocol that outputs a state with the same distillable entanglement as EPL-D and succeeds with larger probability.

We defer the proof of Theorem 3 to Appendix H. The informal argument relies on the fact that the distillable entanglement of the output of a distillation protocol multiplied by the rate of successful distillation cannot be larger than the distillable entanglement of the original state; that is, we must have that

$$p_{\mathrm{succ,EPL}}E_D(\eta_{\hat{A}\hat{B}}(p_d)) \leqslant E_D(\rho_{AB}(p,p_d)). \tag{50}$$

In the case of EPL, the distillable entanglement of the state $\rho_{AB}(p,p_d)$ equals $p_{\mathrm{succ,EPL}}[1-h(p_d)]$ (see Appendix H) while the distillable entanglement of the output state of EPL-D, $\eta_{\hat{A}\hat{B}}(p_d)$, is $1-h(p_d)$, where $h(x)=-x\log x-(1-x)\log(1-x)$ is the binary entropy function [48]. This proves that we actually have equality in Eq. (50). The result is stronger in the case that there is no dephasing, i.e., $p_d=1$. In this case, EPL-D outputs perfect EPR pairs at the distillable entanglement rate. Hence, EPL-D is then by definition optimal within EPL.

E. S states

We have already looked at the R state, a simple mixture of a Bell state with a product state. However, we have only considered the scenario when the product state is orthogonal to the given Bell state. As we have already seen those states are easy to both distill and filter. Specifically, we have seen that from two copies of such a state we can obtain a perfect maximally entangled state with finite probability of success and even from a single copy in the limit of zero probability of success, a perfect maximally entangled state can also be filtered. It is now interesting to see what happens if this product noise is not orthogonal to that Bell state. Hence we will now consider the state

$$\tau_{AB} = p|\Phi^+\rangle\langle\Phi^+| + (1-p)|11\rangle\langle 11|, \tag{51}$$

which we will call an S state.

Numerical examples

The first property of this S state that we have verified numerically is that it is less filterable than the R state, meaning that even at the expense of the probability of success it is not possible to achieve arbitrarily high output fidelity through local filtering. However, we show here that by applying the seesaw optimization from existing schemes to such local filtering of the S state, we find a new protocol that is more suited to those states. Namely, we start from the filtering scheme described in Sec. III A. We see in Fig. 12 that the seesaw method improves the output fidelity of the original filtering protocol designed to perform well on states given in Eq. (10). We observe that the new protocol obtained using the seesaw method overlaps with the PPT bound which proves its optimality for the considered state.

We then investigate distillation on two copies of such an S state. We plot our numerical results in Fig. 13. We see that distilling these states is harder than distilling R states in the sense that the output fidelity of one is no longer achievable for any probability of success. Moreover, our interpolation scheme does not allow for deterministic increase of fidelity which we see is possible using PPT operations. The numerical results also suggest that DEJMPS protocol is optimal for distilling these states, such that it allows us to achieve the highest possible output fidelity for this protocol's probability of success when operating on these states.

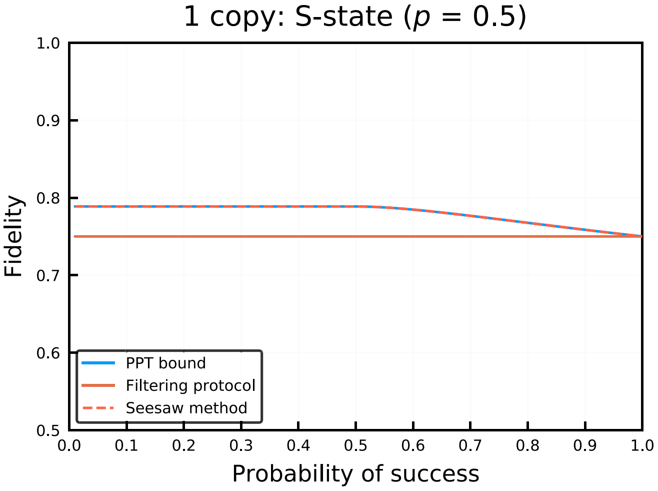

FIG. 12. Filtering S state τ_{AB} with $D = 2$ and $p = 0.5$ in Eq. (51) to a two-qubit state. The fidelity of the input copy is $F_{\text{in}} = 0.75$. We see that deterministic increase of fidelity ($p_{\text{succ}} = 1$) is not possible. We also observe that the filtering scheme designed to work well for states given in Eq. (10) is not able to improve the fidelity of the S state for any value of the probability of success. However, after applying the seesaw method to this protocol we obtain a new filtering protocol that allows for increasing fidelity of this state. Since the curve corresponding to that protocol overlaps with the PPT bound, we see that this protocol is in fact optimal for this state. The 1-BSE bound is not included because it overlaps with the PPT bound.

V. DISCUSSION

We have provided and studied several methods to understand the trade-off between fidelity and probability of success in practical entanglement distillation schemes. The fidelity is thereby of interest not only because it is a commonly estimated measure in experiment, but most significantly because it bears a direct relation to the possible fidelity of teleportation using the entanglement generated [49]. Given that the deterministic transmission of qubits in present day systems relies on the heralded generation of entanglement, followed by deterministic teleportation (see, e.g., [50]), the fidelity is thus of central interest in a quantum network. Evidently, it is an interesting open question to derive trade-offs between the success probability and different entanglement measures.

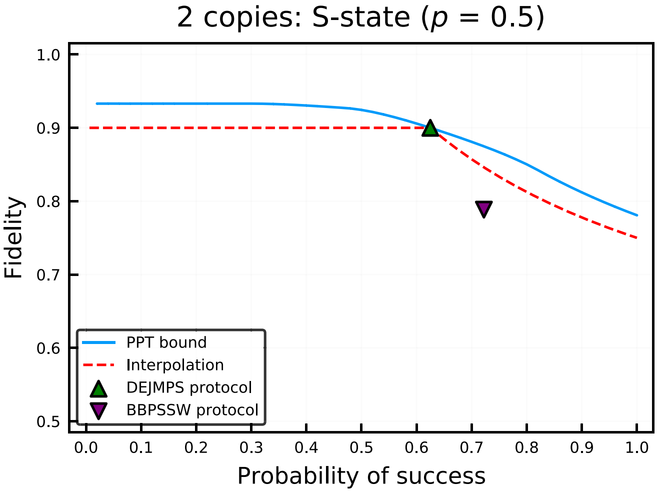

FIG. 13. Distilling the S states $\tau_{ab}^{\otimes 2}$ with $D = 2$ and $p = 0.6$ in Eq. (51) to a two-qubit state. The fidelity of the input copy is $F_{\text{in}} = 0.75$ and we observe that while the extrapolation from DEJMPS does not allow for deterministic distillation (with $p_{\text{succ}} = 1$) in this case, the PPT bound still allows for this possibility. We also observe that DEJMPS allows us to achieve the highest fidelity for the corresponding probability of success. The 1-BSE bound is not included because it overlaps with the PPT bound.

Looking at the method of Bose symmetric extensions employed here, one might wonder whether one might also employ methods based on ϵ nets (see, e.g., [51]) in order to tackle our optimization problem. Here an ϵ net is placed on the set of operations, and every point in this ϵ net is checked. Whereas this "try everything" approach seems rather trivial it does actually (in terms of ϵ) not lead to a computationally (in terms of k) more expensive optimization than the methods of k Bose symmetric extensions when optimizing over the set of separable states. We remark that while this comparison is evidently very interesting and fruitful from a complexity theoretic perspective, it is not of great practical interest for the small values of k for which it is feasible to evaluate the SDP. Here, the corresponding ϵ of the net is still very large, meaning we try out only relatively few points, leading to uninteresting solutions. In contrast, the method of k Bose symmetric extensions actually performs not so badly even for $k = 1$ in a more practical fashion. We remark that the method of ϵ nets can of course be used to optimize over MX operations directly. It is straightforward to adapt the methods of [51] to derive conditions for optimizing over the set of Choi states instead of all states, and then explore the resulting ϵ net to optimize. This evidently leads to statements on the complexity of optimizing over Choi states, but does not lead to a practically realizable method which is the interest of the present article.

One might also wonder whether there exist good heuristic methods based on semidefinite programming in order to derive actual distillation schemes other than using the seesaw method starting from an existing scheme. This indeed may sound quite appealing given heuristics for imposing rank constraints on SDP variables: in our case, we could make explicit a potential ancilla that Alice and Bob may use in their distillation scheme. Fixing an ancilla of a desired maximum size, the Choi state is then pure if we include the purifying ancilla. As such, heuristics such as [52] that confine the rank of the entire state including the ancilla to be 1 approximate the set of pure states, and could thus give rise to a heuristic method for optimizing over MX operations directly. In our situation, however, an implementation of [52] did not lead to any interesting results, which is why this method is omitted from this article. Nevertheless, it is an interesting open question to find good heuristic methods for optimizing over the set of MX operations.

ACKNOWLEDGMENTS

We acknowledge helpful discussions with Stefan Bäuml, Earl Campbell, Ben Criger, Kenneth Goodenough, Peter Humphreys, Jonas Helsen, Karol Horodecki, Jędrzej Kaniewski, Joel Klassen, Victoria Lipinska, Corsin Pfister, Patrick Rall, Jérémy Ribeiro, Bartosz Regula, Liangzhong

Ruan, Mark Steudtner, and Benjamin Vervliet. F.R., T.S., L.P.T., D.E., and S.W. were supported by STW, NWO VIDI, NWO Zwaartekracht QSC, and an ERC Starting Grant. A.D. was supported by the Australian Research Council through the Centre of Excellence in Engineered Quantum Systems (CE110001013 and CE170100009).

F.R. and T.S. contributed equally to this work.

APPENDIX A: PPT CHOI STATES

In this Appendix we briefly discuss the connection between the PPT channels and PPT Choi states. The connection between the PPT channels and Jamiolkowski operator has been discussed in Ref. [21]; however here we are interested in the Choi isomorphism and so for clarity we describe this connection for the Choi isomorphism.

Following [35], we first recall the definition of a PPT operation:

Definition 1. A quantum operation $\Psi_{AB\to\hat{A}\hat{B}}$ is a PPT operation if the superoperator $\Psi^{\Gamma}_{AB\to\hat{A}\hat{B}}$ is completely positive. Here, $\Psi^{\Gamma}_{AB\to\hat{A}\hat{B}}$ is defined such that

$$\Psi^{\Gamma}_{AB\to\hat{A}\hat{B}} : \rho_{AB} \to \left[\Psi_{AB\to\hat{A}\hat{B}}\left(\rho_{AB}^{\Gamma_B}\right)\right]^{\Gamma_{\hat{B}}}, \tag{A1}$$

with Γ_B and $\Gamma_{\hat{B}}$ denoting partial transposes on systems B and $\hat{B}$.

Now we can easily prove that a PPT Choi state corresponds to a PPT operation.

Lemma 1. A quantum operation $\Psi_{AB\to\hat{A}\hat{B}}$ is a PPT operation if and only if its Choi state $C_{\hat{A}\hat{B}A'B'}(\Psi)$ is PPT.

Proof. We use without proof the following simple observation: for every linear map $\Psi_{A\to\hat{A}}$, it follows

$$(\Psi_{A\to\hat{A}} \otimes \mathbb{1}_B)(\Phi_{AB}) = [\mathbb{1}_{\hat{A}} \otimes T_B \circ (\Psi_{\hat{B}\to B})^{\dagger} \circ T_{\hat{B}}](\Phi_{\hat{A}\hat{B}}), \tag{A2}$$

where T denotes the transpose map and $\Psi^{\dagger}$ is the adjoint of Ψ (i.e., the unique linear map satisfying $\mathrm{tr}\,[\rho\Psi(\sigma)] = \mathrm{tr}\,[\sigma\Psi^{\dagger}(\rho)]$).

Consider the Choi matrix of the map Ψ^{Γ}:

$$C_{\hat{A}\hat{B}A'B'}(\Psi^{\Gamma}) = \left(\Psi^{\Gamma}_{AB\to\hat{A}\hat{B}} \otimes \mathbb{1}_{A'B'}\right)\Phi_{ABA'B'} = (T_{\hat{B}} \circ \Psi_{AB\to\hat{A}\hat{B}} \circ T_B \otimes \mathbb{1}_{A'B'})\Phi_{ABA'B'} \tag{A3}$$

$$= [T_{\hat{B}} \circ \Psi_{AB\to\hat{A}\hat{B}} \otimes T_{A'B'} \circ (T_{B'})^{\dagger} \circ T_{A'B'}]\Phi_{ABA'B'}. \tag{A4}$$

It can be easily verified that $(T_{B'})^{\dagger} = T_{B'}$, so that

$$C_{\hat{A}\hat{B}A'B'}(\Psi^{\Gamma}) = (T_{\hat{B}} \otimes T_{B'})C_{\hat{A}\hat{B}A'B'}(\Psi) = [C_{\hat{A}\hat{B}A'B'}(\Psi)]^{\Gamma_{\hat{B}B'}}. \tag{A5}$$

Now it can be clearly seen that

$$[C_{\hat{A}\hat{B}A'B'}(\Psi)]^{\Gamma_{\hat{B}B'}} \geqslant 0 \iff C_{\hat{A}\hat{B}A'B'}(\Psi^{\Gamma}) \geqslant 0 \iff \Psi^{\Gamma} \quad \text{is a completely positive map,} \tag{A6}$$

which concludes the proof. ■

APPENDIX B: BACKGROUND: WELL-KNOWN PROTOCOLS

For convenience, we briefly state the well-known protocols from the literature which we compare to our PPT and 1-BSE bounds. We also describe how we can interpolate or extrapolate new schemes from those existing ones in order to obtain schemes that allow us to succeed with arbitrary desired probability.

1. Fixed protocols

First we state again the filtering protocol [34] that has already been mentioned in Sec. III A:

Algorithm 1 filtering protocol

```
Perform local measurements given by the POVMs: $\{M_A^0, M_A^1\}$ and $\{M_B^0, M_B^1\}$ with $M_A^1 = (A_A^1)^{\dagger}A_A^1$, where $A_A^1 = \sqrt{\epsilon}|0\rangle\langle 0| + |1\rangle\langle 1|$ and $M_A^0 = (A_A^0)^{\dagger}A_A^0 = \mathbb{I} - M_A^1$ and with $M_B^1 = (A_B^1)^{\dagger}A_B^1$, where $A_B^1 = \sqrt{\epsilon}|1\rangle\langle 1| + |0\rangle\langle 0|$ and $M_B^0 = (A_B^0)^{\dagger}A_B^0 = \mathbb{I} - M_B^1$ for some parameter $\epsilon \in [0,1]$.
Communicate the results.
if The measurement outcomes corresponding to $M_A^1$ and $M_B^1$ are obtained then
    Output the post-measurement state.
return
```

This protocol is designed to perform well for the state $\rho_{AB} = p|\Phi_2\rangle\langle\Phi_2| + (1-p)|01\rangle\langle 01|$ defined in Eq. (10) [which is the R state defined in Eq. (42) up to a local bit (and phase) flip]. For this state, conditioned on success the postmeasurement state is $\eta_{\hat{A}\hat{B}} = \frac{p\epsilon}{p_{\mathrm{succ}}}|\Phi_2\rangle\langle\Phi_2| + \frac{(1-p)\epsilon^2}{p_{\mathrm{succ}}}|01\rangle\langle 01|$ with fidelity $F = \frac{p\epsilon}{p_{\mathrm{succ}}}$ and with the probability of success of the filtering procedure given by $p_{\mathrm{succ}} = p\epsilon + (1-p)\epsilon^2$. At the end of Appendix B 2 we describe the modification of this filtering scheme that allows us to achieve higher fidelities for R states with smaller values of p in the case of larger desired probability of success.

Now we will describe the distillation procedures that perform distillation from two to one copies of a two-qubit state. The most generic distillation protocol is the BBPSSW protocol [8] which is applicable to states whose fidelity with some maximally entangled state satisfies $F > 0.5$.

Algorithm 2 BBPSSW protocol

```
Depolarize the two available copies of the state to the isotropic state form:
    $\tau = p|\Phi^+\rangle\langle\Phi^+| + (1-p)\frac{\mathbb{I}}{4},$
 with fidelity $F = (3p+1)/4$.
Apply bilocal CNOT gates between the two copies.
Measure the target qubits and communicate the results.
if The measured flags are 00 or 11 (this occurs with probability $p_{\text{succ}} = F^2 + 2F(1-F)/3 + 5[(1-F)/3]^2$) then
    The source (first) copy becomes more entangled than before (fidelity to $|\Phi^+\rangle$ increases). We obtain a Bell diagonal state with fidelity $F'$ such that
        $F' = \frac{F^2 + [(1-F)/3]^2}{p_{\text{succ}}}.$
return
```

The protocol that can often achieve higher output fidelity than BBPSSW is the DEJMPS protocol [9], which we show is optimal for rank-three Bell diagonal states Eq. (37). Specifically, we consider a version of DEJMPS in which the Bell coefficients are first permuted in a way which maximizes output fidelity [41]. Again, this protocol is applicable to states whose fidelity with some maximally entangled state satisfies $F > 0.5$.

Algorithm 3 DEJMPS protocol

```
Twirl the two available copies of the state to the Bell diagonal state using LOCC.
Perform local rotations on both Alice's and Bob's qubits so that the two copies are in the form
    $\tau = p_1|\Phi^+\rangle\langle\Phi^+| + p_2|\Psi^+\rangle\langle\Psi^+| + p_3|\Phi^-\rangle\langle\Phi^-| + p_4|\Psi^-\rangle\langle\Psi^-|,$
 with $p_1 > 0.5$, $p_1 > p_2 \geqslant p_3 \geqslant p_4$, and $p_1 + p_2 + p_3 + p_4 = 1$. This ordering of the Bell coefficients allows to achieve the highest fidelity [41].
Perform additional rotations: rotate both qubits on Alice's side by $\pi/2$ around the $X$ axis and by $-\pi/2$ on Bob's side.
Apply bilocal CNOT gates between the two copies.
Measure the target qubits and communicate the results.
if The measured flags are 00 or 11 [this occurs with probability $p_{\text{succ}} = (p_1 + p_4)^2 + (p_2 + p_3)^2$] then
    The source (first) copy becomes more entangled than before (fidelity to $|\Phi^+\rangle$ increases). We obtain a state
        $\eta = p_1'|\Phi^+\rangle\langle\Phi^+| + p_2'|\Psi^+\rangle\langle\Psi^+| + p_3'|\Psi^-\rangle\langle\Psi^-| + p_4'|\Phi^-\rangle\langle\Phi^-|,$
    with $p_1' = (p_1^2 + p_4^2)/p_{\text{succ}}$, $p_2' = (p_2^2 + p_3^2)/p_{\text{succ}}$, $p_3' = 2p_2p_3/p_{\text{succ}}$, $p_4' = 2p_1p_4/p_{\text{succ}}$.
return
```

Finally, we also describe the simple protocol first proposed in Ref. [20] which allows us to probabilistically distill a maximally entangled state from two copies of the R state defined in Eq. (42). Since this distillation protocol is utilized within the Extreme Photon Loss (EPL) entanglement generation scheme [7,13] (see below), we refer to it here as EPL-D.

Algorithm 4 EPL-D protocol

```
Apply bilocal CNOT gates between the two copies.
Measure the target qubits and communicate the results.
if The measured flags are 11 then
    Output the source (first) copy.
return
```

When applied to two copies of the R state defined in Eq. (42), the EPL-D protocol extracts a perfect maximally entangled state with probability of success given by $p_{\text{succ}} = p^2/2$. Since R states arise in the remote entanglement generation scheme that uses a single photon detection scheme [44], EPL-D will be a very natural element of such a remote entanglement generation scheme. The scheme for remote entanglement generation using a single photon detection scheme and a distillation operation under the condition of Extreme Photon Loss has been proposed in Ref. [13]. Here we will consider an adaptation of this entanglement generation scheme as proposed in Ref. [7], which performs distillation on the modified version of R states that includes the noise arising from the lack of knowledge about the internal phase of the generated entangled state and possible additional dephasing. The scheme presented in Ref. [7], which we will refer to here as the Extreme Photon Loss (EPL) scheme, utilizes EPL-D to eliminate both the effect of photon loss and lack of knowledge about the internal phase of the generated states. We describe the whole procedure in detail below.

Algorithm 5 EPL entanglement generation scheme

```
1: Generate node-photon entanglement at both remote nodes, where the photonic qubit is encoded in the presence-absence of a photon.
2: Send the photonic qubit towards a beam-splitter station in the middle.
3: Conditioned on the detection of a single photon, store the resulting state in quantum memories.
4: Repeat the above procedure to generate the second copy.
5: Assuming stability of the experimental apparatus over the time of generating those two copies, Alice and Bob share then an effective state
       $\rho_{AB}(p,p_d) = \frac{1}{2\pi}\int d\phi \tau_{A1B1}(\phi,p,p_d) \otimes \tau_{A2B2}(\phi,p,1),$
   where
       $\tau_{AB}(\phi,p,p_d) = p[p_d|\Psi^+(\phi)\rangle\langle\Psi^+(\phi)| + (1-p_d)|\Psi^-(\phi)\rangle\langle\Psi^-(\phi)|] + (1-p)|11\rangle\langle 11|.$
   The dephasing noise corresponds to the decoherence of the quantum memories storing the first copy, while attempting to generate the second one and to the possible small drifts in the phase $\phi$ between the two copies.
6: Apply EPL-D distillation scheme.
7: if EPL-D succeeds (this occurs with probability $p_{\mathrm{succ}} = p^2/2$) then
8:     After Alice applies additional local rotation, we obtain a state
           $\eta_{\hat{A}\hat{B}}(p_d) = p_d|\Phi^+\rangle\langle\Phi^+| + (1-p_d)|\Phi^-\rangle\langle\Phi^-|,$
       with fidelity $p_d$.
9: return
```

2. Interpolating and extrapolating between and from the fixed schemes

We note that having access to shared randomness, Alice and Bob can also apply a mixture of existing schemes. Consider two protocols with probability of success given by p_1 for the first one and p_2 for the second one. Also let the output fidelity conditioned on success be given by F_1 and F_2 for the two protocols, respectively. Then if Alice and Bob share a classical coin with probability distribution $(r,1-r)$, i.e., with probability r the coin outputs head and with probability $1-r$ it outputs tail, then they can construct a new protocol in which they first toss the coin and depending on the outcome they apply either the first or the second scheme. This new scheme has a probability of success given by

$$p_{\mathrm{succ}} = rp_1 + (1-r)p_2, \tag{B1}$$

and the output fidelity conditioned on success will now be given by

$$F = \frac{1}{p_{\mathrm{succ}}}[rp_1F_1 + (1-r)p_2F_2]. \tag{B2}$$

It is also possible to easily extrapolate from an existing scheme. Consider a protocol that succeeds with probability p_1 with the output fidelity conditioned on success given by F_1. Then one can also trivially achieve the same fidelity for any smaller value of p_{succ} by first performing that protocol, then conditioned on its success throwing a coin and effectively accepting the output of the protocol only for one of the outcomes of the coin.

It is also possible to extrapolate in the direction of higher probability of success. For all the considered states apart from the scenario of remote entanglement generation and R states with smaller values of the p parameter, we consider the following extrapolation scheme from a fixed protocol $\mathcal{P}$ when considering distillation from two to one copies. Alice and Bob first throw a coin with probability distribution $(r,1-r)$ and depending on the outcome they either apply the protocol $\mathcal{P}$, which upon success occurring with probability p outputs a state of fidelity F_{out}, or they output one of the input copies of fidelity F_{in}. This scheme has a probability of success

$$p_{\mathrm{succ}} = rp + (1-r), \tag{B3}$$

and the output fidelity conditioned on success will now be given by

$$F = \frac{1}{p_{\mathrm{succ}}}[rpF_{\mathrm{out}} + (1-r)F_{\mathrm{in}}]. \tag{B4}$$

In the case of remote entanglement generation using EPL, the state from which we distill is not a simple tensor product of two copies and therefore the above extrapolation scheme could not be applied in this case. Hence, we then apply a different scheme. In this case Alice and Bob first apply the

EPL-D protocol which upon success occurring with probability p outputs a state of fidelity F_{out}. In the case in which EPL-D fails, they throw a coin with probability distribution $(r,1-r)$. Then for one of the coin outcomes Alice and Bob output a separable state of fidelity $1/2$, and declare failure for the other outcome. This gives

$$p_{\text{succ}} = p + (1-p)r, \tag{B5}$$

with the output fidelity given by

$$F = \frac{1}{p_{\text{succ}}}\left[pF_{\text{out}} + (1-p)r\,\frac{1}{2}\right]. \tag{B6}$$

It also turns out that for R states with $F_{\text{in}} < 2-\sqrt{2}$ it is also better in terms of output fidelities to apply this extrapolation scheme to EPL-D without interpolating with DEJMPS at all.

Finally we also describe the extrapolation-based modified filtering protocol which we apply to the states defined in Eq. (10) (rotated R states). In this scheme Alice and Bob apply the filtering protocol as described in Appendix B 1, but in the case of failure they throw a coin with probability distribution $(r,1-r)$ and depending on the outcome they either output a state of fidelity half or declare a failure. This leads to the new overall probability of success given by $p_{\text{succ}} = p\epsilon + (1-p)\epsilon^2 + [1 - p\epsilon - (1-p)\epsilon^2]r$ and new output fidelity given by $F = \{2p\epsilon + [1 - p\epsilon - (1-p)\epsilon^2]r\}/2p_{\text{succ}}$. For fixed value of the probability of success one can then optimize the fidelity over ϵ and r. The result shows that the modification (throwing a coin with nonzero probability of outputting a product state) helps for $p < 2/3$ for larger values of the success probability. In particular after fixing p_{succ} the optimal output fidelity that can be obtained using this protocol is given by

$$F = \begin{cases} \frac{1}{2}\left(1 + \frac{p^2}{4p_{\text{succ}}(1-p)}\right), & p \leqslant \frac{2}{3} \wedge p_{\text{succ}} \geqslant \frac{3p^2}{4(1-p)}, \\ \frac{2p}{p+\sqrt{p^2+4p_{\text{succ}}(1-p)}}, & \text{otherwise}. \end{cases} \tag{B7}$$

We note that it is the first parameter regime of the above function where probabilistically adding the product noise of fidelity half helps. The second regime corresponds to just applying the original filtering scheme. We also note that setting $p_{\text{succ}} = 1$ in the above expression we recover the result of [47] for maximum fidelity obtainable from a single copy of the R state using trace preserving LOCC operations.

APPENDIX C: SYMMETRY REDUCTION

If the structure of the SDP optimization exhibits a certain symmetry we can exploit this to simplify the optimization before actually evaluating it numerically. Inspired by the observation of Rains [21] we make a similar symmetry reduction to the main SDP in this section. Specifically, note that the target maximally entangled state Φ_D satisfies

$$\forall U, \quad U_{\hat{A}} \otimes U^*_{\hat{B}}(\Phi_D)(U_{\hat{A}} \otimes U^*_{\hat{B}})^\dagger = \Phi_D. \tag{C1}$$

Let $\mathcal{T}(\cdot)$ be the twirling operation defined as

$$\mathcal{T}(\rho_{\hat{A}\hat{B}}) = \int dU(U_{\hat{A}} \otimes U^*_{\hat{B}})\rho_{AB}(U_{\hat{A}} \otimes U^*_{\hat{B}})^\dagger. \tag{C2}$$

We can then reexpress the symmetry in Eq. (C1) as $\mathcal{T}(\Phi_D) = \Phi_D$. This means that without loss of generality our optimal solution exhibits the same symmetry, because both the constraints and objective function of the SDP in Optimization Program 4 are invariant under the symmetry:

$$\begin{aligned} \text{objective}: \frac{|A||B|}{\delta}\text{tr}\big(|\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}}\otimes\rho^T_{A'B'}\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}\big) &= \frac{|A||B|}{\delta}\text{tr}\big\{\big[\mathcal{T}(|\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}})\otimes\rho^T_{A'B'}\big]\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}\big\} \\ &= \frac{|A||B|}{\delta}\text{tr}\big[\big(|\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}}\otimes\rho^T_{A'B'}\big)\mathcal{T}^\dagger(\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'})\big] = \frac{|A||B|}{\delta}\text{tr}\big[\big(|\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}}\otimes\rho^T_{A'B'}\big)\mathcal{T}(\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'})\big], \\ \text{constraints}: |A||B|\text{tr}\big[\big(\mathbb{I}_{\hat{A}\hat{B}}\otimes\rho^T_{A'B'}\big)\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}\big] &= |A||B|\text{tr}\big[\big(\mathbb{I}_{\hat{A}\hat{B}}\otimes\rho^T_{A'B'}\big)\mathcal{T}(\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'})\big], \end{aligned} \tag{C3}$$

and similarly for the other constraints. In other words, if $\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}$ is an optimal solution, then so is

$$\mathcal{T}(\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}) = \int dU(U_{\hat{A}} \otimes U^*_{\hat{B}} \otimes \mathbb{I}_{A'B'})\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}(U_{\hat{A}} \otimes U^*_{\hat{B}} \otimes \mathbb{I}_{A'B'})^\dagger, \tag{C4}$$

and it is intuitive that $\mathcal{T}(\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'})$ contains a smaller number of variables compared to $\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}$. Thus, declaring and optimizing over the variable $\mathcal{T}(\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'})$ is a more efficient approach.

In order to explicitly write down the symmetry-reduced optimization, we need to understand the structure of the twirling operation (C2). Using the tools from representation theory of the unitary group [53] we can write

$$\mathcal{T}(\rho_{\hat{A}\hat{B}}) = \text{tr}_{\hat{A}\hat{B}}[\rho_{\hat{A}\hat{B}}|\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}}]|\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}} + \text{tr}_{\hat{A}\hat{B}}[\rho_{\hat{A}\hat{B}}(\mathbb{I} - |\Phi_D\rangle\langle\Phi_D|)_{\hat{A}\hat{B}}]\frac{\mathbb{I}_{\hat{A}\hat{B}} - |\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}}}{D^2-1}. \tag{C5}$$

This gives us the new form of our optimization variable as follows:

$$\begin{aligned} \mathcal{T}(\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}) = {}& \text{tr}_{\hat{A}\hat{B}}[\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}(|\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}} \otimes \mathbb{I}_{A'B'})] \otimes |\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}} + \text{tr}_{\hat{A}\hat{B}}[\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'} \\ & \times ((\mathbb{I} - |\Phi_D\rangle\langle\Phi_D|)_{\hat{A}\hat{B}} \otimes \mathbb{I}_{A'B'})] \otimes \frac{\mathbb{I}_{\hat{A}\hat{B}} - |\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}}}{D^2-1}. \end{aligned} \tag{C6}$$

With the definitions

$$M_{A'B'} := \mathrm{tr}_{\hat{A}\hat{B}}[\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}(|\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}} \otimes \mathbb{I}_{A'B'})], \tag{C7}$$

$$E_{A'B'} := \mathrm{tr}_{\hat{A}\hat{B}}[\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}((\mathbb{I} - |\Phi_D\rangle\langle\Phi_D|)_{\hat{A}\hat{B}} \otimes \mathbb{I}_{A'B'})], \tag{C8}$$

we have

$$\mathcal{T}(\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'}) = M_{A'B'} \otimes |\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}} + E_{A'B'} \otimes \frac{\mathbb{I}_{\hat{A}\hat{B}} - |\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}}}{D^2 - 1}, \tag{C9}$$

and it is evident that we have reduced the number of variables to those contained in $M_{A'B'}$ and $E_{A'B'}$.

Now we are ready to derive the form of our SDP in terms of the new variables $M_{A'B'}$ and $E_{A'B'}$. Using (C9) in the objective function gives

$$\frac{|A||B|}{\delta}\,\mathrm{tr}\left[\left(|\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}} \otimes \rho^T_{A'B'}\right)\mathcal{T}(\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'})\right] = \frac{|A||B|}{\delta}\,\mathrm{tr}\left[\rho^T_{A'B'} M_{A'B'}\right]. \tag{C10}$$

Similarly, the equality constraint transforms as

$$|A||B|\,\mathrm{tr}\left[\left(\mathbb{I}_{\hat{A}\hat{B}} \otimes \rho^T_{A'B'}\right)\mathcal{T}(\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'})\right] = |A||B|\,\mathrm{tr}\left[\rho^T_{A'B'}(M_{A'B'} + E_{A'B'})\right] = \delta. \tag{C11}$$

The inequality constraint $\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'} \geqslant 0$ becomes two inequality constraints $M_{A'B'} \geqslant 0$ and $E_{A'B'} \geqslant 0$. The PPT relaxation constraint $\hat{C}^{\Gamma}_{\checkmark,\hat{A}A'\hat{B}B'} \geqslant 0$ becomes

$$\begin{aligned}
\mathcal{T}(\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'})^{\Gamma} &= |\Phi_D\rangle\langle\Phi_D|^{\Gamma}_{\hat{A}\hat{B}} \otimes M^{\Gamma}_{A'B'} + \frac{(\mathbb{I}_{\hat{A}\hat{B}} - |\Phi_D\rangle\langle\Phi_D|_{\hat{A}\hat{B}})^{\Gamma}}{D^2 - 1} \otimes E^{\Gamma}_{A'B'} \\
&= \frac{1}{D}(P_{S_{\hat{A}\hat{B}}} - P_{A_{\hat{A}\hat{B}}}) \otimes M^{\Gamma}_{A'B'} + \frac{\left(1 - \frac{1}{D}\right)P_{S_{\hat{A}\hat{B}}} + \left(1 + \frac{1}{D}\right)P_{A_{\hat{A}\hat{B}}}}{D^2 - 1} \otimes E^{\Gamma}_{A'B'} \\
&= P_{S_{\hat{A}\hat{B}}} \otimes \left(\frac{1}{D}M^{\Gamma}_{A'B'} + \frac{1 - \frac{1}{D}}{D^2 - 1}E^{\Gamma}_{A'B'}\right) + P_{A_{\hat{A}\hat{B}}} \otimes \left(-\frac{1}{D}M^{\Gamma}_{A'B'} + \frac{1 + \frac{1}{D}}{D^2 - 1}E^{\Gamma}_{A'B'}\right) \geqslant 0,
\end{aligned} \tag{C12}$$

where we have used $\Phi^{\Gamma} = (P_S - P_A)/D$ and $\mathbb{I}^{\Gamma} = P_S + P_A$, where P_S and P_A are projectors onto the symmetric and antisymmetric subspaces, respectively. The orthogonality of P_S and P_A allows us to read off this constraint as two inequality constraints

$$M^{\Gamma}_{A'B'} + \frac{1}{D+1}E^{\Gamma}_{A'B'} \geqslant 0, \quad -M^{\Gamma}_{A'B'} + \frac{1}{D-1}E^{\Gamma}_{A'B'} \geqslant 0. \tag{C13}$$

Finally, the last two inequality constraints of SDP in Optimization Program 4 become

$$M_{A'B'} + E_{A'B'} = \mathrm{tr}_{\hat{A},\hat{B}}[\mathcal{T}(\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'})] = \hat{C}_{\checkmark,A'B'} \leqslant \frac{\mathbb{I}_{A',B'}}{|A||B|}, \tag{C14}$$

$$M^{\Gamma}_{A'B'} + E^{\Gamma}_{A'B'} = \{\mathrm{tr}_{\hat{A},\hat{B}}[\mathcal{T}(\hat{C}_{\checkmark,\hat{A}A'\hat{B}B'})]\}^{\Gamma} = \hat{C}^{\Gamma}_{\checkmark,A'B'} \leqslant \frac{\mathbb{I}_{A',B'}}{|A||B|}. \tag{C15}$$

In summary, putting things together we obtain the following simplified SDP optimization problem, as stated in Optimization Program 5 in the main text:

$$\begin{aligned}
\text{maximize} \quad & \frac{|A||B|}{\delta}\,\mathrm{tr}\left[\rho^T_{A'B'} M_{A'B'}\right] \\
\text{subject to} \quad & M_{A'B'} \geqslant 0,\, E_{A'B'} \geqslant 0, \\
& M_{A'B'} + E_{A'B'} \leqslant \frac{\mathbb{I}_{A'B'}}{|A||B|}, \\
& M^{\Gamma}_{A'B'} + E^{\Gamma}_{A'B'} \leqslant \frac{\mathbb{I}_{A'B'}}{|A||B|}, \\
& |A||B|\,\mathrm{tr}\left[\rho^T_{A'B'}(M_{A'B'} + E_{A'B'})\right] = \delta, \\
& M^{\Gamma}_{A'B'} + \frac{1}{D+1}E^{\Gamma}_{A'B'} \geqslant 0, \\
& -M^{\Gamma}_{A'B'} + \frac{1}{D-1}E^{\Gamma}_{A'B'} \geqslant 0.
\end{aligned}$$

Optimization Program 7.

APPENDIX D: DERIVATIONS OF DUAL SDPs

In this Appendix we will restate some results of the theory of semidefinite programming, particularly the dual SDP, following the approach of Watrous [54]. We will use these results to derive the form of the dual SDPs for optimizing fidelity and probability of success.

There are various ways of presenting a general semidefinite program. It is most convenient for our purposes to use the following form, given in Ref. [54], for an SDP and its dual:

(1) Primal:

$$\begin{aligned} \text{maximize} \quad & \operatorname{tr}[AX] \\ \text{subject to} \quad & \Phi_1(X) = B_1, \\ & \Phi_2(X) \leqslant B_2, \\ & X \geqslant 0. \end{aligned}$$

Optimization Program 8.

(2) Dual:

$$\begin{aligned} \text{minimize} \quad & \operatorname{tr}[B_1 Y_1] + \operatorname{tr}[B_2 Y_2] \\ \text{subject to} \quad & \Phi_1^\dagger(Y_1) + \Phi_2^\dagger(Y_2) \geqslant A, \\ & Y_1 = Y_1^\dagger, \\ & Y_2 \geqslant 0. \end{aligned}$$

Optimization Program 9.

Here A, B_1, B_2 are Hermitian matrices, Φ_1 and Φ_2 are Hermiticity preserving linear maps and $\Phi^\dagger$ is a Hermiticity preserving linear map uniquely defined in terms of Φ through the following relation: $\operatorname{tr}[\Phi(X)Y] = \operatorname{tr}[X\Phi^\dagger(Y)]$ for all Hermitian matrices X and Y. Notice that the map $\Phi^\dagger$ reverses the order of the spaces as compared to the original map Φ. The variables of the primal SDP are the matrix elements of the Hermitian matrix X and any X that satisfies the constraints is termed a *feasible* X. Likewise the variables of the dual SDP are the Hermitian matrices Y_1 and Y_2, and such matrices are termed feasible if they satisfy the constraints of the dual SDP. It is a very straightforward observation that feasible points of the dual SDP can be used to provide bounds on the primal optimum and vice versa. To show this consider feasible variables X, Y_1, Y_2; then we have

$$\begin{aligned} &\operatorname{tr}[B_1 Y_1] + \operatorname{tr}[B_2 Y_2] - \operatorname{tr}[AX] \\ &\quad = \operatorname{tr}[\Phi_1(X) Y_1] + \operatorname{tr}[\Phi_2(X) Y_2] \\ &\qquad + \operatorname{tr}\{[B_2 - \Phi_2(X)] Y_2\} - \operatorname{tr}[AX] \\ &\quad = \operatorname{tr}\{X[\Phi_1^\dagger(Y_1) + \Phi_2^\dagger(Y_2) - A]\} \\ &\qquad + \operatorname{tr}\{[B_2 - \Phi_2(X)] Y_2\} \geqslant 0. \end{aligned} \tag{D1}$$

The first equality just comes from implementing the equality constraints of the primal SDP. The second equality is just an application of the definition of $\Phi^\dagger$, and the final inequality arises from the inequality constraints of the SDP and the fact that $\operatorname{tr}[XY] \geqslant 0$ if $X \geqslant 0$ and $Y \geqslant 0$. This observation is known as *weak duality* and, as stated in the main text, it is the key tool that we will use to provide bounds on the single-shot distillation fidelity with fixed probability of success.

1. Optimizing fidelity

The SDP in Optimization Program 5 for finding the optimal output fidelity can be written in the above form by defining

$$\begin{aligned} &A = \frac{|A||B|}{\delta}\begin{pmatrix} \rho^T & 0 \\ 0 & 0 \end{pmatrix}, \quad X = \begin{pmatrix} M & X_{12} \\ X_{12}^\dagger & E \end{pmatrix}, \quad B_1 = \delta, \quad B_2 = \begin{pmatrix} \frac{\mathbb{I}}{|A||B|} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\mathbb{I}}{|A||B|} \end{pmatrix}, \\ &\Phi_1(X) = |A||B| \operatorname{tr}[\rho^T(M+E)], \\ &\Phi_2(X) = \begin{pmatrix} M+E & 0 & 0 & 0 \\ 0 & -M^\Gamma - \frac{1}{D+1}E^\Gamma & 0 & 0 \\ 0 & 0 & M^\Gamma - \frac{1}{D-1}E^\Gamma & 0 \\ 0 & 0 & 0 & M^\Gamma + E^\Gamma \end{pmatrix}. \end{aligned} \tag{D2}$$

Observe that the SDP induced by this choice is equivalent to the original SDP in Optimization Program 5 because the constraint $X \geqslant 0$ reduces to $M \geqslant 0$ and $E \geqslant 0$ without loss of generality. More precisely, the $X \geqslant 0$ implies $M \geqslant 0$ and $E \geqslant 0$ so the optimum of the original SDP in Optimization Program 5 is at least as large as the optimum of the SDP defined here. Conversely, for any feasible pair M, E of the original SDP in Optimization Program 5 we can define a feasible X of the above SDP by setting $X_{12} = 0$ so the optimum of the original SDP in Optimization Program 5 is at most the optimum of the above SDP.

Now in order to dualize, we need to calculate $\Phi_1^\dagger$ and $\Phi_2^\dagger$. Since Φ_1 maps to a scalar, we conclude that $Y_1 = y$ is a scalar and we must have, by definition of adjoint,

$$\operatorname{tr}[\Phi_1(X) Y_1] = |A||B| \operatorname{tr}[\rho^T(M+E)] y = \operatorname{tr}[X\Phi_1^\dagger(Y_1)], \tag{D3}$$

from which we conclude that

$$\Phi_1^\dagger(Y_1) = |A||B| \begin{pmatrix} \rho^T y & 0 \\ 0 & \rho^T y \end{pmatrix}. \tag{D4}$$

Turning now to Φ_2, we note that Y_2 will be a 4×4 block matrix and we will label the blocks as Y_2^{ij}. Observe that

$$\begin{aligned}\mathrm{tr}[\Phi_2(X)Y_2] &= \mathrm{tr}\left[(M+E)Y_2^{11}\right] + \mathrm{tr}\left[\left(-M^\Gamma - \frac{1}{D+1}E^\Gamma\right)Y_2^{22}\right] + \mathrm{tr}\left[\left(M^\Gamma - \frac{1}{D-1}E^\Gamma\right)Y_2^{33}\right] + \mathrm{tr}\left[(M^\Gamma + E^\Gamma)Y_2^{44}\right]\\ &= \mathrm{tr}\left[(M+E)Y_2^{11}\right] + \mathrm{tr}\left[\left(-M - \frac{1}{D+1}E\right)\left(Y_2^{22}\right)^\Gamma\right] + \mathrm{tr}\left[\left(M - \frac{1}{D-1}E\right)\left(Y_2^{33}\right)^\Gamma\right] + \mathrm{tr}\left[(M+E)\left(Y_2^{44}\right)^\Gamma\right].\end{aligned} \tag{D5}$$

With $\Phi_2^\dagger(Y_2)$ expressed as a 2×2 block matrix

$$\Phi_2^\dagger(Y_2) = \begin{pmatrix} W_1 & W_2 \\ W_2^\dagger & W_4 \end{pmatrix}, \tag{D6}$$

we have

$$\mathrm{tr}[X\Phi_2^\dagger(Y_2)] = \mathrm{tr}[MW_1] + \mathrm{tr}[X_{12}^\dagger W_2] + \mathrm{tr}[X_{12}W_2^\dagger] + \mathrm{tr}[EW_4]. \tag{D7}$$

The definition of the adjoint map, namely $\mathrm{tr}\,[\Phi_2(X)Y_2] = \mathrm{tr}\,[X\Phi_2^\dagger(Y_2)]$, allows us to directly compare (D5) and (D7) and read off

$$W_1 = Y_2^{11} - \left(Y_2^{22}\right)^\Gamma + \left(Y_2^{33}\right)^\Gamma + \left(Y_2^{44}\right)^\Gamma, \quad W_2 = 0, \quad W_3 = 0, \quad W_4 = Y_2^{11} - \frac{1}{D+1}\left(Y_2^{22}\right)^\Gamma - \frac{1}{D-1}\left(Y_2^{33}\right)^\Gamma + \left(Y_2^{44}\right)^\Gamma. \tag{D8}$$

Therefore the dual program becomes

$$\begin{aligned}\text{minimize}\quad & y\delta + \frac{\mathrm{tr}\left[Y_2^{11} + Y_2^{44}\right]}{|A||B|}\\ \text{subject to}\quad & \begin{pmatrix} |A||B|y\rho^T + Y_2^{11} - \left(Y_2^{22}\right)^\Gamma + \left(Y_2^{33}\right)^\Gamma + \left(Y_2^{44}\right)^\Gamma & 0 \\ 0 & |A||B|y\rho^T + Y_2^{11} - \frac{1}{D+1}\left(Y_2^{22}\right)^\Gamma - \frac{1}{D-1}\left(Y_2^{33}\right)^\Gamma + \left(Y_2^{44}\right)^\Gamma \end{pmatrix}\\ & \geqslant \begin{pmatrix} \frac{|A||B|}{\delta}\rho & 0 \\ 0 & 0 \end{pmatrix},\\ & y \in \mathbb{R},\\ & Y_2 \geqslant 0.\end{aligned}$$

Optimisation Program 10.

For ease of notation we will define $J = Y_2^{11}, G = Y_2^{22}, H = Y_2^{33}, K = Y_2^{44}$. The off-diagonal blocks of the matrix variable Y_2 can always be chosen to be zero and thus the dual SDP can be written as follows without loss of generality:

$$\begin{aligned}\text{minimize}\quad & y\delta + \frac{\mathrm{tr}[J+K]}{|A||B|}\\ \text{subject to}\quad & J, G, H, K \geqslant 0, y \in \mathbb{R},\\ & |A||B|\left(y - \frac{1}{\delta}\right)\rho^T + J - G^\Gamma + H^\Gamma + K^\Gamma \geqslant 0,\\ & |A||B|y\rho^T + J - \frac{1}{D+1}G^\Gamma - \frac{1}{D-1}H^\Gamma + K^\Gamma \geqslant 0.\end{aligned}$$

Optimisation Program 11.

Here all the matrices are on registers $A'B'$. Thus we have obtained the form of the dual semidefinite program for the optimal output fidelity.

2. Optimizing probability of success

Similarly, we can now find the dual of the SDP in Optimization Program 6 for optimizing probability of success. Again, using the form specified in Ref. [54], we obtain

$$A = |A||B|\begin{pmatrix} \rho^T & 0 \\ 0 & \rho^T \end{pmatrix}, \quad X = \begin{pmatrix} M & X_{12} \\ X_{12}^{\dagger} & E \end{pmatrix}, \quad B_1 = 0, \quad B_2 = \begin{pmatrix} \frac{\mathbb{I}}{|A||B|} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\mathbb{I}}{|A||B|} \end{pmatrix},$$

$$\Phi_1(X) = (1-F)\,\mathrm{tr}[\rho^T M] - F\,\mathrm{tr}[\rho^T E], \tag{D9}$$

$$\Phi_2(X) = \begin{pmatrix} M+E & 0 & 0 & 0 \\ 0 & -M^{\Gamma} - \frac{1}{D+1}E^{\Gamma} & 0 & 0 \\ 0 & 0 & M^{\Gamma} - \frac{1}{D-1}E^{\Gamma} & 0 \\ 0 & 0 & 0 & M^{\Gamma} + E^{\Gamma} \end{pmatrix}.$$

Now we need to calculate $\Phi_1^{\dagger}$ and $\Phi_2^{\dagger}$. Since Φ_1 maps to a scalar, we conclude that $Y_1 = y$ is a scalar and we must have, by definition of adjoint,

$$\mathrm{tr}[\Phi_1(X), Y_1] = \{(1-F)\,\mathrm{tr}[\rho^T M] - F\,\mathrm{tr}[\rho^T E]\}y = \mathrm{tr}[X\Phi_1^{\dagger}(Y_1)], \tag{D10}$$

from which we conclude that

$$\Phi_1^{\dagger}(Y_1) = \begin{pmatrix} (1-F)y\rho^T & 0 \\ 0 & -Fy\rho^T \end{pmatrix}. \tag{D11}$$

Turning now to Φ_2, we note that it is the same as in the program for optimizing fidelity; see Eq. (D2). Hence $\Phi_2^{\dagger}(Y_2)$ remains the same as given in Eq. (D6) and in Eq. (D8).

Therefore the dual problem becomes

$$\begin{aligned} \text{minimize} \quad & \frac{\mathrm{tr}\left[Y_2^{11} + Y_2^{44}\right]}{|A||B|} \\ \text{subject to} \quad & \begin{pmatrix} (1-F)y\rho^T + Y_2^{11} - \left(Y_2^{22}\right)^{\Gamma} + \left(Y_2^{33}\right)^{\Gamma} + \left(Y_2^{44}\right)^{\Gamma} & 0 \\ 0 & -Fy\rho^T + Y_2^{11} - \frac{1}{D+1}\left(Y_2^{22}\right)^{\Gamma} - \frac{1}{D-1}\left(Y_2^{33}\right)^{\Gamma} + \left(Y_2^{44}\right)^{\Gamma} \end{pmatrix} \\ & \geqslant |A||B|\begin{pmatrix} \rho^T & 0 \\ 0 & \rho^T \end{pmatrix}, \\ & y \in \mathbb{R}, \\ & Y_2 \geqslant 0. \end{aligned}$$

Optimisation Program 12.

This SDP can be rewritten as

$$\begin{aligned} \text{minimize} \quad & \frac{\mathrm{tr}[J+K]}{|A||B|} \\ \text{subject to} \quad & J, G, H, K \geqslant 0, y \in \mathbb{R}, \\ & [(1-F)y - |A||B|]\rho^T + J - G^{\Gamma} + H^{\Gamma} + K^{\Gamma} \geqslant 0, \\ & [-Fy - |A||B|]\rho^T + J - \frac{1}{D+1}G^{\Gamma} - \frac{1}{D-1}H^{\Gamma} + K^{\Gamma} \geqslant 0. \end{aligned}$$

Optimisation Program 13.

APPENDIX E: *k* BOSE SYMMETRIC EXTENSIONS

This section details the calculations leading to the 1-BSE optimization program mentioned in the main text. We first explain how the variable is defined for a k-BSE. Considering $\hat{C}_{(\hat{A}A')\hat{B}B'}$ to be k-BSE means that there exists $\hat{C}_{(\hat{A}_1A'_1)\ldots(\hat{A}_{k+1}A'_{k+1})\hat{B}B'}$ satisfying the BSE constraints. We are changing the optimization variable from the former to the latter, which lives only on the symmetric subspace of $(\hat{A}_1A'_1)\ldots(\hat{A}_{k+1}A'_{k+1})$. The full Hilbert space of Alice decomposes as

$$\mathcal{H}_{(\hat{A}_1A'_1)\ldots(\hat{A}_{k+1}A'_{k+1})} = \mathcal{H}_{\mathrm{Sym}} \oplus \mathcal{H}_{\mathrm{Sym}}^{\perp}, \tag{E1}$$

into symmetric subspace and its orthogonal complement. Hence, the joint Hilbert space of Alice's and Bob's systems has the corresponding form

$$\mathcal{H}_{(\hat{A}_1A'_1)\dots(\hat{A}_{k+1}A'_{k+1})\hat{B}B'} = (\mathcal{H}_{\text{Sym}} \oplus \mathcal{H}_{\text{Sym}}^{\perp}) \otimes \mathcal{H}_{\hat{B},B'} = (\mathcal{H}_{\text{Sym}} \otimes \mathcal{H}_{\hat{B},B'}) \oplus (\mathcal{H}_{\text{Sym}}^{\perp} \otimes \mathcal{H}_{\hat{B},B'}). \tag{E2}$$

Under this decomposition, the operator $\hat{C}_{(\hat{A}_1A'_1)\dots(\hat{A}_{k+1}A'_{k+1})\hat{B}B'}$ has the simple form

$$\hat{C}_{(\hat{A}_1A'_1)\dots(\hat{A}_{k+1}A'_{k+1})\hat{B}B'} = \begin{pmatrix} W_s & 0 \\ 0 & 0 \end{pmatrix}, \tag{E3}$$

with W_s being some operator acting on $\mathcal{H}_{\text{Sym}} \otimes \mathcal{H}_{\hat{B},B'}$. Since our derivations in the main text are performed in the standard basis, let $U_{\text{Sym}\to\text{Std}}$ be the change of basis from the "symmetric" basis to the computational basis of Alice's systems. We finally obtain the form of our new variable in the standard basis

$$\hat{C}_{(\hat{A}_1A'_1)\dots(\hat{A}_{k+1}A'_{k+1})\hat{B}B'} = U_{\text{Sym}\to\text{Std}} \otimes \mathbb{I}_{\hat{B},B'} \begin{pmatrix} W_s & 0 \\ 0 & 0 \end{pmatrix} U_{\text{Sym}\to\text{Std}}^{\dagger} \otimes \mathbb{I}_{\hat{B},B'}. \tag{E4}$$

In the final SDP which will be presented at the end of this section, we will only declare and optimize over the smaller variable W_s.

Now we specialize to the case of 1-BSE. Considering $\hat{C}_{(\hat{A}A')\hat{B}B'}$ to be 1-BSE means that there exists $\hat{C}_{(\hat{A}_1A'_1)(\hat{A}_2A'_2)\hat{B}B'}$ satisfying the BSE constraints. Since we have only two subsystems on Alice's side (corresponding to the indices 1 and 2), the orthogonal complement $\mathcal{H}_{\text{Sym}}^{\perp}$ turns out to be the subspace consisting of antisymmetric vectors $\mathcal{H}_{\text{ASym}}$. We need to compute the change of basis operator in

$$\hat{C}_{\hat{A}_1A'_1\hat{A}_2A'_2\hat{B}B'} = U_{\text{Sym}\to\text{Std}} \otimes \mathbb{I}_{\hat{B},B'} \begin{pmatrix} W_s & 0 \\ 0 & 0 \end{pmatrix} U_{\text{Sym}\to\text{Std}}^{\dagger} \otimes \mathbb{I}_{\hat{B},B'}. \tag{E5}$$

In the case when the input dimensions of Alice and Bob are the same and the target is the maximally entangled state of dimension D, we have dimensions $|\hat{A}_1| = |\hat{A}_2| = |\hat{B}| = D$ and $|A'_1| = |A'_2| = |B'| = C$, so Alice's first $(\hat{A}_1A'_1)$ and second $(\hat{A}_2A'_2)$ subsystems each have dimension CD. We can construct the change of basis $U_{\text{Sym}\to\text{Std}}$ for $\mathbb{C}^{CD} \otimes \mathbb{C}^{CD}$ using standard techniques. Let $\{|i\rangle : i = 0, \dots, CD\}$ denote the standard basis of a CD-dimensional system. Then the basis for the symmetric subspace on $(A_1A'_1)(A_2A'_2)$ consists of the vectors in $V_s = V_1 \cup V_2$ where

$$\begin{aligned} V_1 &= \big\{|i\rangle_{A_1A'_1} \otimes |i\rangle_{A_2A'_2} | i = 0,1,\dots,CD\big\}, \\ V_2 &= \left\{\frac{1}{\sqrt{2}}\big(|i\rangle_{A_1A'_1} \otimes |j\rangle_{A_2A'_2} + |j\rangle_{A_1A'_1} \otimes |i\rangle_{A_2A'_2}\big) \Big| i,j = 0,1,\dots,CD \text{ and } j > i\right\}. \end{aligned} \tag{E6}$$

Similarly, the basis for the antisymmetric subspace on $(A_1A'_1)(A_2A'_2)$ consists of the vectors in

$$V_a = \left\{\frac{1}{\sqrt{2}}\big(|i\rangle_{A_1A'_1} \otimes |j\rangle_{A_2A'_2} - |j\rangle_{A_1A'_1} \otimes |i\rangle_{A_2A'_2}\big) \Big| i,j = 0,1,\dots,CD \text{ and } j > i\right\}. \tag{E7}$$

The coefficients of these vectors form the entries of the matrix $U_{\text{Sym}\to\text{Std}}$.

We are now left with rewriting the optimization in terms of W_s, a $\frac{(CD)^2(CD+1)}{2} \times \frac{(CD)^2(CD+1)}{2}$ matrix. The objective function

$$\frac{|A||B|}{\delta} \operatorname{tr}\left\{\big(\mathbb{I}_{\hat{A}_1A'_1} \otimes |\Phi_D\rangle\langle\Phi_D|_{\hat{A}_2,\hat{B}} \otimes \rho^T_{A'_2B'}\big)\left[U_{\text{Sym}\to\text{Std}} \otimes \mathbb{I}_{\hat{B},B'} \begin{pmatrix} W_s & 0 \\ 0 & 0 \end{pmatrix} U_{\text{Sym}\to\text{Std}}^{\dagger} \otimes \mathbb{I}_{\hat{B}B'}\right]\right\} \tag{E8}$$

can be rewritten as (since the trace is cyclic under permutation of operators)

$$\operatorname{tr}\left[X \begin{pmatrix} W_s & 0 \\ 0 & 0 \end{pmatrix}\right], \tag{E9}$$

where we convert the input data written in standard basis to the "symmetric" basis

$$X = \frac{|A||B|}{\delta} U_{\text{Sym}\to\text{Std}}^{\dagger} \otimes \mathbb{I}_{\hat{B}B'} \big(\mathbb{I}_{\hat{A}_1A'_1} \otimes |\Phi_D\rangle\langle\Phi_D|_{\hat{A}_2\hat{B}} \otimes \rho^T_{A'_2B'}\big) U_{\text{Sym}\to\text{Std}} \otimes \mathbb{I}_{\hat{B}B'}. \tag{E10}$$

This means that only X_s, the component of X living in the symmetric subspace, i.e., the first $\frac{(CD)^2(CD+1)}{2}$ rows and columns of X, will appear in the objective function and the objective function becomes $\operatorname{tr}(X_sW_s)$. Similarly, the constraint on the probability of success can be rewritten as $\operatorname{tr}(Y_sW_s) = \delta$, where

$$Y = |A||B| U_{\text{Sym}\to\text{Std}}^{\dagger} \otimes \mathbb{I}_{\hat{B}B'} \big(\mathbb{I}_{\hat{A}_1A'_1} \otimes \mathbb{I}_{\hat{A}_2\hat{B}} \otimes \rho^T_{A'_2B'}\big) U_{\text{Sym}\to\text{Std}} \otimes \mathbb{I}_{\hat{B}B'}, \tag{E11}$$

and again Y_s is just a matrix that consists of the first $\frac{(CD)^2(CD+1)}{2}$ rows and columns of Y. All other constraints become unaffected so the SDP becomes

$$\begin{aligned}
\text{maximize} \quad & \operatorname{tr}\left(X_{s\,\hat{A}_1A'_1\hat{A}_2A'_2\hat{B}B'}W_{s\,\hat{A}_1A'_1\hat{A}_2,A'_2,\hat{B},B'}\right)\\
\text{subject to} \quad & \operatorname{tr}\left(Y_{s\,\hat{A}_1A'_1\hat{A}_2A'_2\hat{B}B'}W_{s\,\hat{A}_1A'_1\hat{A}_2A'_2\hat{B}B'}\right)=\delta,\\
& W_{s\,\hat{A}_1A'_1\hat{A}_2A'_2\hat{B}B'}\geqslant 0,\\
& \operatorname{tr}_{\hat{A}_1A'_1}\left[U_{\text{Sym}\to\text{Std}}\otimes\mathbb{I}_{\hat{B},B'}\begin{pmatrix}W_s & 0\\ 0 & 0\end{pmatrix}U^{\dagger}_{\text{Sym}\to\text{Std}}\otimes\mathbb{I}_{\hat{B},B'}\right]^{\Gamma}\geqslant 0,\\
& \operatorname{tr}_{\hat{A}_1A'_1\hat{A}_2\hat{B}}\left[U_{\text{Sym}\to\text{Std}}\otimes\mathbb{I}_{\hat{B}B'}\begin{pmatrix}W_s & 0\\ 0 & 0\end{pmatrix}U^{\dagger}_{\text{Sym}\to\text{Std}}\otimes\mathbb{I}_{\hat{B},B'}\right]\leqslant\frac{\mathbb{I}_{A'_2B'}}{|A||B|},\\
& \operatorname{tr}_{\hat{A}_1A'_1\hat{A}_2\hat{B}}\left[U_{\text{Sym}\to\text{Std}}\otimes\mathbb{I}_{\hat{B}B'}\begin{pmatrix}W_s & 0\\ 0 & 0\end{pmatrix}U^{\dagger}_{\text{Sym}\to\text{Std}}\otimes\mathbb{I}_{\hat{B},B'}\right]^{\Gamma}\leqslant\frac{\mathbb{I}_{A'_2B'}}{|A||B|}.
\end{aligned}$$

Optimization Program 14.

In the scenario most frequently considered in this paper, that is of distillation from two to one copies of a two-qubit state, we have that $C = 4$ and $D = 2$ and so our variable W_s is a 288×288 matrix.

APPENDIX F: DEFINITIONS OF OPTIMALITY

In this section we introduce certain terminology that will later allow us to make precise optimality claims of the different distillation protocols. We also introduce and prove specific lemmas that later allow us to prove our optimality claims with respect to the EPL-D protocol in Appendix H.

Let Λ denote the map corresponding to a distillation protocol and $P_{\checkmark}$ be the projector on the success space of the flags. We introduce the following shorthands:

$$\Psi(\Lambda,P_{\checkmark},\rho)=\operatorname{tr}_F[(\mathbb{I}_{\hat{A}\hat{B}}\otimes P_{\checkmark})\Lambda_{AB\to\hat{A}\hat{B}F}(\rho)], \tag{F1}$$

$$\eta(\Lambda,P_{\checkmark},\rho)=\frac{\Psi(\Lambda,P_{\checkmark},\rho)}{p(\Lambda,P_{\checkmark},\rho)}, \tag{F2}$$

where

$$p(\Lambda,P_{\checkmark},\rho)=\operatorname{tr}[\Psi(\Lambda,P_{\checkmark},\rho)]. \tag{F3}$$

That is, Ψ,η are, respectively, the unnormalized and normalized output state conditioned on success. We introduce two additional shorthands for the fidelity of Ψ and η to $|\Phi^+\rangle = |\Phi_2\rangle$, which for simplicity we will now denote as simply Φ:

$$g(\Lambda,P_{\checkmark},\rho)=F(\Psi(\Lambda,P_{\checkmark},\rho),\Phi), \tag{F4}$$

$$f(\Lambda,P_{\checkmark},\rho)=F(\eta(\Lambda,P_{\checkmark},\rho),\Phi). \tag{F5}$$

Note that $\eta(\Lambda,P_{\checkmark},\rho)$ and $f(\Lambda,P_{\checkmark},\rho)$ are defined only if $p(\Lambda,P_{\checkmark},\rho) > 0$.

We define the optimal output fidelity $f_{\text{opt}}(\rho)$ and the optimal success probability $p_{\text{opt}}(\rho)$ when optimized over all LOCC distillation operations Λ and success projectors $P_{\checkmark}$ as follows:

$$f_{\text{opt}}(\rho)=\sup_{\Lambda\in\text{LOCC},\,P_{\checkmark}|p(\Lambda,P_{\checkmark},\rho)>0} f(\Lambda,P_{\checkmark},\rho) \tag{F6}$$

and

$$p_{\text{opt}}(\rho)=\sup_{\Lambda\in\text{LOCC},\,P_{\checkmark}|p(\Lambda,P_{\checkmark},\rho)>0\text{ and }f(\Lambda,P_{\checkmark},\rho)=f_{\text{opt}}(\rho)} p(\Lambda,P_{\checkmark},\rho). \tag{F7}$$

With this notation, we introduce two different definitions of optimality:

Definition 2. We call a protocol Λ with the success projector $P_{\checkmark}$ *fidelity-optimal* with respect to the quantum state ρ if

$$f(\Lambda,P_{\checkmark},\rho)=f_{\text{opt}}(\rho) \tag{F8}$$

and

$$p(\Lambda,P_{\checkmark},\rho)=p_{\text{opt}}(\rho). \tag{F9}$$

We emphasize here that the above definition concerns distillation towards the maximally entangled state with $D = 2$, but it can be easily generalized to higher values of D.

Definition 3. We call a protocol Λ with the success projector $P_{\checkmark}$ *distillation-optimal* with respect to the quantum state ρ if

$$p(\Lambda,P_{\checkmark},\rho)E_D(\eta(\Lambda,P_{\checkmark},\rho)) = E_D(\rho), \tag{F10}$$

where $E_D(\rho)$ is the distillable entanglement of ρ.

Note that our definition of a protocol being distillation-optimal implies that no protocol can achieve a better trade-off between success probability and distillable entanglement of the output state (Lemma 4). We recall that the distillable entanglement is defined as an optimization over arbitrary distillation protocols and, in general, can only be achieved if Alice and Bob hold an infinite number of copies of the state ρ.

In the following, we prove several basic facts of these definitions.

Lemma 2. Let $\rho = \sum_i \lambda_i \rho_i$ such that $\forall i,\lambda_i > 0$ and $\sum_i \lambda_i = 1$. Then,

$$f_{\text{opt}}\left(\sum_i \lambda_i \rho_i\right) \leqslant \max_i f_{\text{opt}}(\rho_i). \tag{F11}$$

Proof.

$$\begin{aligned} f_{\text{opt}}\left(\sum_i \lambda_i \rho_i\right) &= \sup_{\Lambda\in\text{LOCC},P_{\checkmark}|p(\Lambda,P_{\checkmark},\rho)>0} \frac{g(\Lambda,P_{\checkmark},\sum_i \lambda_i\rho_i)}{p(\Lambda,P_{\checkmark},\sum_j \lambda_j\rho_j)} = \sup_{\Lambda\in\text{LOCC},P_{\checkmark}|p(\Lambda,P_{\checkmark},\rho)>0} \frac{\sum_{i|p(\Lambda,P_{\checkmark},\rho_i)>0} \lambda_i f(\Lambda,P_{\checkmark},\rho_i)p(\Lambda,P_{\checkmark},\rho_i)}{\sum_j \lambda_j p(\Lambda,P_{\checkmark},\rho_j)} \\ &\leqslant \max_i f_{\text{opt}}(\rho_i). \end{aligned} \tag{F12}$$

■

Lemma 3. Let $\rho = \sum_i \lambda_i \rho_i$ such that $\forall i,\lambda_i > 0$ and $\sum_i \lambda_i = 1$, let Λ and $P_{\checkmark}$ correspond to a distillation protocol such that $f(\Lambda,P_{\checkmark},\rho) = f_{\text{opt}}(\rho) = \max_i f_{\text{opt}}(\rho_i)$, and let the index k be such that $f(\Lambda,P_{\checkmark},\rho_k) = \max_i f(\Lambda,P_{\checkmark},\rho_i)$ is unique. Then,

$$p(\Lambda,P_{\checkmark},\rho) \leqslant \lambda_k. \tag{F13}$$

Proof. From Lemma 2 we see that we must have

$$f(\Lambda,P_{\checkmark},\rho_k) = f_{\text{opt}}(\rho) = \max_i f_{\text{opt}}(\rho_i) = f_{\text{opt}}(\rho_k). \tag{F14}$$

Then,

$$\begin{aligned} f_{\text{opt}}(\rho_k) &= f\left(\Lambda,P_{\checkmark},\sum_i \lambda_i\rho_i\right) \\ &= \frac{\sum_{i|p(\Lambda,P_{\checkmark},\rho_i)>0} \lambda_i f(\Lambda,P_{\checkmark},\rho_i)p(\Lambda,P_{\checkmark},\rho_i)}{p(\Lambda,P_{\checkmark},\rho)} \\ &= \frac{\lambda_k p(\Lambda,P_{\checkmark},\rho_k)}{p(\Lambda,P_{\checkmark},\rho)} f_{\text{opt}}(\rho_k) + \sum_{i \neq kp(\Lambda,P_{\checkmark},\rho_i) > 0} \frac{\lambda_i p(\Lambda,P_{\checkmark},\rho_i)}{p(\Lambda,P_{\checkmark},\rho)} f(\Lambda,P_{\checkmark},\rho_i). \end{aligned} \tag{F15}$$

Now note that $\sum_i \lambda_i p(\Lambda,P_{\checkmark},\rho_i)/p(\Lambda,P_{\checkmark},\rho) = 1$ and $\forall i \neq k, f(\Lambda,P_{\checkmark},\rho_i) < f_{\text{opt}}(\rho_k)$. That is, we have a convex combination of $f_{\text{opt}}(\rho_k)$ and all the other $f(\Lambda,P_{\checkmark},\rho_i)$ that are smaller than $f_{\text{opt}}(\rho_k)$. As this convex combination needs to equal $f_{\text{opt}}(\rho_k)$, we require that $\frac{\lambda_k p(\Lambda,P_{\checkmark},\rho_k)}{p(\Lambda,P_{\checkmark},\rho)} = 1$ and $\forall i \neq k, p(\Lambda,P_{\checkmark},\rho_i) = 0$. This means that

$$p(\Lambda,P_{\checkmark},\rho) = \lambda_k p(\Lambda,P_{\checkmark},\rho_k) \leqslant \lambda_k. \tag{F16}$$

■

Lemma 4. Given a bipartite state ρ and an LOCC protocol $\Lambda_{AB\to\hat{A}\hat{B}F}$ together with a projector $P_{\checkmark}$, it holds that

$$p(\Lambda,P_{\checkmark},\rho)E_D(\eta(\Lambda,P_{\checkmark},\rho)) \leqslant E_D(\rho). \tag{F17}$$

Proof. Suppose that there exists $\Lambda_{AB\to\hat{A}\hat{B}F}$ together with a projector $P_{\checkmark}$ such that

$$p(\Lambda,P_{\checkmark},\rho)E_D(\eta(\Lambda,P_{\checkmark},\rho)) > E_D(\rho). \tag{F18}$$

Then it would be possible to take n copies of ρ, obtain approximately $np(\Lambda,P_{\checkmark},\rho)$ copies of $\eta(\Lambda,P_{\checkmark},\rho)$, and for large enough n distill $np(\Lambda,P_{\checkmark},\rho)E_D(\eta(\Lambda,P_{\checkmark},\rho))$ EPR pairs which would be strictly larger than $nE_D(\rho)$. However, this is not possible since by definition $E_D(\rho)$ is the maximum rate at which EPR pairs can be distilled from ρ_{AB} by LOCC. ■

APPENDIX G: BELL DIAGONAL STATES

In Sec. IV B 2, we stated Theorem 1 and argued that the DEJMPS distillation protocol is optimal for distilling two copies of rank-three Bell diagonal states. In this appendix we make this argument rigorous. The formal statement that we show is as follows:

Theorem 4. DEJMPS is fidelity-optimal with respect to the state $\rho = \tau^{\otimes 2}$, where

$$\tau = p_1|\Phi^+\rangle\langle\Phi^+| + p_2|\Psi^+\rangle\langle\Psi^+| + (1-p_1-p_2)|\Phi^-\rangle\langle\Phi^-|, \tag{G1}$$

with $p_1 > 0.5$ and $p_1 > p_2 \geqslant 1 - p_1 - p_2$.

Remark 1. Every Bell diagonal state of rank up to three can be transformed to the form in Eq. (G1) using only local Clifford operations; hence Theorem 4 effectively applies to all Bell diagonal states of rank up to three.

The proof is structured as follows. In Appendix G 1, we prove some basic properties of Bell diagonal states. In Appendix G 2, we show that DEJMPS protocol achieves $f(\text{DEJMPS},\rho) = f_{\text{opt}}(\rho)$ for states of the form in Eq. (G1) and we complete the argument in Appendix G 3, where we show that the success probability for these states is $p(\text{DEJMPS},\rho) = p_{\text{opt}}(\rho)$.

1. Properties of the Bell diagonal states

Consider the Bell diagonal states

$$\tau = p_1|\Phi^+\rangle\langle\Phi^+| + p_2|\Psi^+\rangle\langle\Psi^+| + p_3|\Phi^-\rangle\langle\Phi^-| + (1-p_1-p_2-p_3)|\Psi^-\rangle\langle\Psi^-|. \tag{G2}$$

Given the parameters (p_1,p_2,p_3) we have that $\text{tr}\,[\tau] = 1$ and the eigenvalues of τ are positive so long as $p_1,p_2,p_3 \geqslant 0$ and $1 - p_1 - p_2 - p_3 \geqslant 0$. Geometrically the set of Bell diagonal states forms a tetrahedron. Notice that $p_1 = \text{tr}\,[|\Phi^+\rangle\langle\Phi^+|\tau]$ and so on.

We can give an alternative parametrization for τ as follows:

$$\tau = \tfrac{1}{4}(II + r_1XX + r_2YY + r_3ZZ), \tag{G3}$$

where for Pauli matrices P_i we use the shorthand notation $P_i \otimes P_j = P_iP_j$. Notice that $r_1 = \text{tr}\,[XX\tau]$ and so on. The convenience of this parametrization is that

$$\tau^\Gamma = \tfrac{1}{4}(II + r_1XX - r_2YY + r_3ZZ), \tag{G4}$$

so that in these coordinates the partial transpose is a reflection. (This follows because $Y^T = -Y$ and other Pauli matrices are unaffected by transpose.) Notice that the partial transpose of a Bell diagonal state is a Bell diagonal matrix.

We can use the definitions to find

$$p_1 = (1 + r_1 - r_2 + r_3)/4, \tag{G5}$$

$$p_2 = (1 + r_1 + r_2 - r_3)/4, \tag{G6}$$

$$p_3 = (1 - r_1 + r_2 + r_3)/4, \tag{G7}$$

$$1 - p_1 - p_2 - p_3 = (1 - r_1 - r_2 - r_3)/4. \tag{G8}$$

These formulas make it possible to tell when τ is positive even if it is expressed in terms of the parameters r_i. Now if we have two copies of τ we of course have

$$\begin{aligned}\tau \otimes \tau &= \tfrac{1}{4}(II + r_1XX + r_2YY + r_3ZZ) \otimes \tfrac{1}{4}(II + r_1XX + r_2YY + r_3ZZ)\\ &= \tfrac{1}{16}\big[IIII + r_1(IIXX + XXII) + r_2(IIYY + YYII) + r_3(IIZZ + ZZII) + r_1^2XXXX\\ &\quad + r_1r_2(XXYY + YYXX) + r_1r_3(XXZZ + ZZXX) + r_2^2YYYY + r_2r_3(YYZZ + ZZYY)\big].\end{aligned} \tag{G9}$$

In the dual SDP we will restrict attention to dual variables V that have the same symmetry as the matrices $\tau \otimes \tau$; specifically,

$$\begin{aligned}V = \tfrac{1}{16}[&v_0IIII + v_1(IIXX + XXII) + v_2(IIYY + YYII) + v_3(IIZZ + ZZII) + v_{11}XXXX + v_{12}(XXYY + YYXX)\\ &+ v_{13}(XXZZ + ZZXX) + v_{22}YYYY + v_{23}(YYZZ + ZZYY)]\end{aligned} \tag{G10}$$

and so

$$\begin{aligned}V^\Gamma = \tfrac{1}{16}[&v_0IIII + v_1(IIXX + XXII) - v_2(IIYY + YYII) + v_3(IIZZ + ZZII) + v_{11}XXXX - v_{12}(XXYY + YYXX)\\ &+ v_{13}(XXZZ + ZZXX) + v_{22}YYYY - v_{23}(YYZZ + ZZYY)].\end{aligned} \tag{G11}$$

Here Γ denotes the transpose on Bob's systems, that is, on the second and fourth Pauli matrices. Notice that in this parametrization $v_{13} = \mathrm{tr}\,[(XXZZ)V]$ and so on. Alternatively we can expand V in terms of projections on the Bell states as follows:

$$\begin{aligned} V = {} & w_1|\Phi^+\rangle\langle\Phi^+||\Phi^+\rangle\langle\Phi^+| + w_2(|\Phi^+\rangle\langle\Phi^+||\Psi^+\rangle\langle\Psi^+| + |\Psi^+\rangle\langle\Psi^+||\Phi^+\rangle\langle\Phi^+|) + w_3|\Psi^+\rangle\langle\Psi^+||\Psi^+\rangle\langle\Psi^+| \\ & + w_4|\Phi^-\rangle\langle\Phi^-||\Phi^-\rangle\langle\Phi^-| + w_5(|\Phi^+\rangle\langle\Phi^+||\Phi^-\rangle\langle\Phi^-| + |\Phi^-\rangle\langle\Phi^-||\Phi^+\rangle\langle\Phi^+|) + w_6(|\Psi^+\rangle\langle\Psi^+||\Phi^-\rangle\langle\Phi^-| \\ & + |\Phi^-\rangle\langle\Phi^-||\Psi^+\rangle\langle\Psi^+|) + w_7|\Psi^-\rangle\langle\Psi^-||\Psi^-\rangle\langle\Psi^-| + w_8(|\Phi^+\rangle\langle\Phi^+||\Psi^-\rangle\langle\Psi^-| + |\Psi^-\rangle\langle\Psi^-||\Phi^+\rangle\langle\Phi^+|) \\ & + w_9(|\Psi^+\rangle\langle\Psi^+||\Psi^-\rangle\langle\Psi^-| + |\Psi^-\rangle\langle\Psi^-||\Psi^+\rangle\langle\Psi^+|) + w_{10}(|\Phi^-\rangle\langle\Phi^-||\Psi^-\rangle\langle\Psi^-| + |\Psi^-\rangle\langle\Psi^-||\Phi^-\rangle\langle\Phi^-|). \end{aligned} \tag{G12}$$

Here we use a shorthand notation $|\psi\rangle\langle\psi| \otimes |\phi\rangle\langle\phi| = |\psi\rangle\langle\psi||\phi\rangle\langle\phi|$. In terms of this parametrization $V \geqslant 0$ if and only if $w_i \geqslant 0$ for all i.

In constructing a dual semidefinite program in the main text we consider a restricted set of V such that $V^\Gamma = V$. It is clear from Eqs. (G10) and (G11) that the condition $V^\Gamma = V$ is equivalent to $v_2 = 0 = v_{12} = v_{23}$. Thus we require the following three conditions:

$$v_2 = -w_1 + w_3 + w_4 + 2w_6 - w_7 - 2w_8 = 0, \tag{G13}$$

$$v_{12} = -w_1 + w_3 - w_4 + 2w_5 + w_7 - 2w_9 = 0, \tag{G14}$$

$$v_{23} = -w_1 + 2w_2 - w_3 + w_4 + w_7 - 2w_{10} = 0. \tag{G15}$$

In the main text we construct a dual feasible solution for the SDP that arises in the restricted case of a Bell diagonal state where $1 - p_1 - p_2 - p_3 = 0$, and therefore $p_3 = 1 - p_1 - p_2$. Making the definitions

$$\lambda_1 = p_1^2,\ \lambda_2 = p_1 p_2,\ \lambda_3 = p_2^2,\ \lambda_4 = (1 - p_1 - p_2)^2,\ \lambda_5 = p_1(1 - p_1 - p_2),\ \lambda_6 = p_2(1 - p_1 - p_2), \tag{G16}$$

we obtain

$$\begin{aligned} \tau \otimes \tau = {} & \lambda_1|\Phi^+\rangle\langle\Phi^+||\Phi^+\rangle\langle\Phi^+| + \lambda_2(|\Phi^+\rangle\langle\Phi^+||\Psi^+\rangle\langle\Psi^+| + |\Psi^+\rangle\langle\Psi^+||\Phi^+\rangle\langle\Phi^+|) + \lambda_3|\Psi^+\rangle\langle\Psi^+||\Psi^+\rangle\langle\Psi^+| \\ & + \lambda_4|\Phi^-\rangle\langle\Phi^-||\Phi^-\rangle\langle\Phi^-| + \lambda_5(|\Phi^+\rangle\langle\Phi^+||\Phi^-\rangle\langle\Phi^-| + |\Phi^-\rangle\langle\Phi^-||\Phi^+\rangle\langle\Phi^+|) \\ & + \lambda_6(|\Psi^+\rangle\langle\Psi^+||\Phi^-\rangle\langle\Phi^-| + |\Phi^-\rangle\langle\Phi^-||\Psi^+\rangle\langle\Psi^+|). \end{aligned} \tag{G17}$$

2. Optimal fidelity of DEJMPS

We will first show that $f(\text{DEJMPS},\rho) = f_{\text{opt}}(\rho)$, when ρ consists of two copies of some Bell diagonal state of rank up to three. The dual SDP for maximizing fidelity has the form

$$\begin{aligned} \text{minimize}\quad & d(y,J,G,H,K) = y\delta + \frac{\mathrm{tr}[J+K]}{|A||B|} \\ \text{subject to}\quad & J,G,H,K \geqslant 0, y \in \mathbb{R}, \\ & |A||B|\left(y - \frac{1}{\delta}\right)\rho^T + J - G^\Gamma + H^\Gamma + K^\Gamma \geqslant 0, \\ & |A||B|y\rho^T + J - \frac{1}{D+1}G^\Gamma - \frac{1}{D-1}H^\Gamma + K^\Gamma \geqslant 0. \end{aligned}$$

Optimization Program 15.

For rank-two and rank-three Bell diagonal states, the output fidelity of DEJMPS is $F = p_1' = p_1^2/N$, where $N = p_1^2 + (1 - p_1)^2$ is the probability that the protocol succeeds. Hence we require a feasible solution of the dual program whose objective function takes the value p_1'. Here we find such a solution that is valid for all $\delta \in (0,1]$. As an ansatz consider a solution with $y = \frac{p_1'}{\delta}$ and $J = G = K = 0$. This means that the objective function takes the value p_1'. We now need to show that there exists a matrix H such that

$$H \geqslant 0, \tag{G18}$$

$$\frac{|A||B|}{\delta}(p_1' - 1)\rho^T + H^\Gamma \geqslant 0, \tag{G19}$$

$$\frac{|A||B|}{\delta}p_1'\rho^T - H^\Gamma \geqslant 0. \tag{G20}$$

To make it simpler we can assume that $H = \frac{|A||B|}{\delta} V$ and so now we need to find the matrix V such that

$$V \geqslant 0, \tag{G21}$$

$$(p_1' - 1)\rho^T + V^\Gamma \geqslant 0, \tag{G22}$$

$$p_1'\rho^T - V^\Gamma \geqslant 0. \tag{G23}$$

Since the input state in our SDP is $\rho = \tau \otimes \tau$ given by Eq. (G17), we further restrict V by requiring that $V = V^\Gamma$. We can also ignore the transpose on $\rho_{A'B'}$ in the above equations as here we work with the Bell diagonal states. The chosen dual variable V that satisfies the above conditions can be specified as follows in terms of the coefficients in Eq. (G12):

$$\begin{gathered} w_1 = p_1'(1-p_1)^2, \quad w_2 = p_1'(1-p_1)p_2, \quad w_3 = p_1'p_2^2, \quad w_4 = p_1'(1-p_1-p_2)^2, \\ w_5 = p_1'(1-p_1)(1-p_1-p_2), \quad w_6 = p_1'p_2(1-p_1-p_2), \quad w_7 = 0 = w_8 = w_9 = w_{10}. \end{gathered} \tag{G24}$$

Clearly $V \geqslant 0$ since $w_i \geqslant 0$ for all i. It is straightforward to check that each of equations (G13)–(G15) are satisfied and therefore $V = V^\Gamma$. Since V^Γ is diagonal in the same basis as $\rho_{A'B'}$, to verify the conditions (G22) and (G23) we just need to verify a set of scalar equations:

$$(p_1' - 1)\lambda_i + w_i \geqslant 0, \tag{G25}$$

$$p_1'\lambda_i - w_i \geqslant 0, \tag{G26}$$

where the coefficients λ_i are defined in Eq. (G16). It is straightforward to determine that each of these equations is satisfied so long as $p_1 \geqslant 1/2$ as was specified originally. This shows that V defined through Eqs. (G12) and (G24) satisfies Eqs. (G22) and (G23) and therefore we have found a feasible solution of the dual problem for which the objective function takes the value p_1' for all values of $\delta \in (0,1]$. This proves that for all those values of δ there exists no protocol that can achieve higher fidelity than p_1', and hence DEJMPS protocol achieves the highest fidelity for two copies of all Bell diagonal states of rank up to three, when optimizing over all LOCC protocols.

3. Optimal probability of success of DEJMPS

Now we will show that DEJMPS also satisfies the second condition required for being fidelity-optimal, namely $p(\text{DEJMPS},\rho) = p_{\text{opt}}(\rho)$. In other words, we will show that it is also not possible to achieve the output fidelity of DEJMPS with probability of success larger than that of DEJMPS. We recall that the dual SDP for the probability of success reads

$$\begin{aligned} \text{minimize} \quad & \frac{\text{tr}[J+K]}{|A||B|} \\ \text{subject to} \quad & J,G,H,K \geqslant 0, \quad y \in \mathbb{R}, \\ & [(1-F)y - |A||B|]\rho^T + J - G^\Gamma + H^\Gamma + K^\Gamma \geqslant 0, \\ & [-Fy - |A||B|]\rho^T + J - \frac{1}{D+1}G^\Gamma - \frac{1}{D-1}H^\Gamma + K^\Gamma \geqslant 0. \end{aligned}$$

Optimisation Program 16.

As an ansatz we consider a solution with $J = |A||B|R$, $y = |A||B|s$ and $G = K = 0$, where

$$\begin{aligned} R = \big[& p_1^2|\Phi^+\rangle\langle\Phi^+||\Phi^+\rangle\langle\Phi^+| + p_2^2|\Psi^+\rangle\langle\Psi^+||\Psi^+\rangle\langle\Psi^+| + (1-p_1-p_2)^2|\Phi^-\rangle\langle\Phi^-||\Phi^-\rangle\langle\Phi^-| \\ & + p_2(1-p_1-p_2)(|\Psi^+\rangle\langle\Psi^+||\Phi^-\rangle\langle\Phi^-| + |\Phi^-\rangle\langle\Phi^-||\Psi^+\rangle\langle\Psi^+|)\big] \end{aligned} \tag{G27}$$

and

$$s = -\frac{N}{(1-p_1)(2p_1-1)}. \tag{G28}$$

This means that the objective function takes the value N. We now need to show that there exists a matrix H such that

$$H \geqslant 0, \tag{G29}$$

$$[(1-F)y - |A||B|]\rho^T + J + H^\Gamma \geqslant 0, \tag{G30}$$

$$[-Fy - |A||B|]\rho^T + J - \frac{1}{D-1}H^\Gamma \geqslant 0. \tag{G31}$$

To make it simpler we can assume that $H = |A||B|V$ and so now we need to find the matrix V such that

$$V \geqslant 0, \tag{G32}$$

$$[(1-F)s-1]\rho^T + R + V^\Gamma \geqslant 0, \tag{G33}$$

$$[-Fs-1]\rho^T + R - \frac{1}{D-1}V^\Gamma \geqslant 0. \tag{G34}$$

Here $F = p_1'$ is the output fidelity of DEJMPS and $N = p_1^2 + (1-p_1)^2$. Again, since we work in the Bell basis with Bell diagonal states, we can ignore the transpose in the above equations. We specify the Bell coefficients of V as

$$w_1 = \frac{(1-p_1)p_1^2}{2p_1-1}, \quad w_2 = \frac{p_1^2 p_2}{2p_1-1}, \quad w_3 = \frac{p_1^2 p_2^2}{(1-p_1)(2p_1-1)}, \quad w_4 = \frac{p_1^2(1-p_1-p_2)^2}{(1-p_1)(2p_1-1)},$$
$$w_5 = \frac{p_1^2(1-p_1-p_2)}{2p_1-1}, \quad w_6 = \frac{p_1^2 p_2(1-p_1-p_2)}{(1-p_1)(2p_1-1)}, \quad w_7 = w_8 = w_9 = w_{10} = 0, \tag{G35}$$

where w's are the Bell coefficients as expressed in the definition Eq. (G12). Now we will show that these variables satisfy all the constraints. Clearly $V \geqslant 0$ since $w_i \geqslant 0$ for all i. It is straightforward to check that each of equations (G13)–(G15) are satisfied and therefore $V = V^\Gamma$. Since V^Γ is diagonal in the same basis as $\rho_{A'B'}$, to verify the conditions (G22) and (G23) we just need to verify a set of scalar equations:

$$[(1-F)s-1]\lambda_i + [R]_{ii} + w_i \geqslant 0, \tag{G36}$$

$$(-Fs-1)\lambda_i + [R]_{ii} - w_i \geqslant 0, \tag{G37}$$

where the coefficients λ_i are again defined in Eq. (G16) and $[R]_{ii}$ are the diagonal entries of R in the Bell basis.

We can easily check that for $p_1 > 0.5$, all the constraints are satisfied and so we have found a feasible solution to the dual SDP for probability of success. The value of the objective function is $\frac{\mathrm{tr}[J]}{|A||B|} = N$. Hence we have found a feasible solution of the dual minimization problem (that provides upper bounds for achievable probability of success) that can be in fact achieved with DEJMPS. That is, we have proven that DEJMPS is also optimal with respect to probability of success. That is, for Bell diagonal states of rank up to three, it is impossible to achieve the output fidelity of DEJMPS with probability of success larger than that of DEJMPS. This concludes the proof that DEJMPS is fidelity-optimal for two copies of all Bell diagonal states of rank up to three.

APPENDIX H: REMOTE ENTANGLEMENT GENERATION THROUGH EPL SCHEME

Here, we show that EPL-D is the optimal distillation protocol within the EPL remote entanglement generation scheme according to our two definitions. That is, we formally state and prove Theorems 2 and 3 which we now formulate as one theorem:

Theorem 5. EPL-D is both fidelity-optimal and distillation-optimal for states of the form

$$\rho_{AB}(p,p_d) = \frac{1}{2\pi}\int d\phi\, \tau_{A1B1}(\phi,p,p_d) \otimes \tau_{A2B2}(\phi,p,1), \tag{H1}$$

where

$$\tau_{AB}(\phi,p,p_d) = p[p_d|\Psi^+(\phi)\rangle\langle\Psi^+(\phi)| + (1-p_d)|\Psi^-(\phi)\rangle\langle\Psi^-(\phi)|] + (1-p)|11\rangle\langle 11|. \tag{H2}$$

We postpone the proof of fidelity-optimal to Appendix H 1 and the proof of distillation-optimal to Appendix H 2.

1. EPL-D is fidelity-optimal

We note that for states of the form Eq. (47) the integration over the phase can be performed analytically:

$$\rho_{AB}(p,p_d) = \frac{p^2}{4}\big[P_{\mathrm{odd}_{A1B1}} \otimes P_{\mathrm{odd}_{A2B2}} + (2p_d-1)(|01\rangle\langle 10|_{A1B1} \otimes |10\rangle\langle 01|_{A2B2} + |10\rangle\langle 01|_{A1B1} \otimes |01\rangle\langle 10|_{A2B2})\big]$$
$$+ \frac{(1-p)p}{2}\big[|11\rangle\langle 11|_{A1B1} \otimes P_{\mathrm{odd}_{A2B2}} + P_{\mathrm{odd}_{A1B1}} \otimes |11\rangle\langle 11|_{A2B2}\big] + (1-p)^2|11\rangle\langle 11|_{A1B1} \otimes |11\rangle\langle 11|_{A2B2}, \tag{H3}$$

where $P_{\text{odd}} = |01\rangle\langle 01| + |10\rangle\langle 10|$ is the projector on the odd-parity subspace of the two-qubit space. Let us now permute the order of the registers to $A1A2B1B2$. After the permutation, ρ takes the following diagonal form in the standard basis:

$$\rho_{AB}(p,p_d) = \begin{pmatrix} 0_3 &&&&&&&&& \\ & a &&&&&&&& \\ && 0_2 &&&&&&& \\ &&& Q &&&&&& \\ &&&& 0_1 &&&&& \\ &&&&& b &&&& \\ &&&&&& a &&& \\ &&&&&&& b && \\ &&&&&&&& b & \\ &&&&&&&&& c \end{pmatrix}, \tag{H4}$$

where 0_i denotes an $i \times i$ zero matrix, all the nonfilled elements are 0, and the shorthands Q,a,b,c, and d stand for

$$Q = \begin{pmatrix} a & 0 & 0 & ad \\ 0 & b & 0 & 0 \\ 0 & 0 & 0 & 0 \\ ad & 0 & 0 & a \end{pmatrix}, \tag{H5}$$

$$a = \frac{p^2}{4}, \tag{H6}$$

$$b = \frac{1}{2}(1-p)p, \tag{H7}$$

$$c = (1-p)^2, \tag{H8}$$

$$d = 2p_d - 1. \tag{H9}$$

Let

$$L(p,p_d) = \begin{pmatrix} 0_3 &&&&&&& \\ & a &&&&&& \\ && 0_7 &&&&& \\ &&& b &&&& \\ &&&& a &&& \\ &&&&& b && \\ &&&&&& b & \\ &&&&&&& c \end{pmatrix}, \quad I(p,p_d) = \begin{pmatrix} 0_6 &&&&& \\ & 0 & 0 & 0 & 0 & \\ & 0 & b & 0 & 0 & \\ & 0 & 0 & 0 & 0 & \\ & 0 & 0 & 0 & 0 & \\ &&&&& 0_6 \end{pmatrix},$$

$$F(p,p_d) = \begin{pmatrix} 0_6 &&&&& \\ & a & 0 & 0 & ad & \\ & 0 & 0 & 0 & 0 & \\ & 0 & 0 & 0 & 0 & \\ & ad & 0 & 0 & a & \\ &&&&& 0_6 \end{pmatrix}. \tag{H10}$$

Now we can rewrite ρ as a function of L,I, and F:

$$\rho_{AB}(p,p_d) = \text{tr}[L]\rho^L + \text{tr}[I]\rho^I + \text{tr}[F]\rho^F, \tag{H11}$$

where

$$\rho^L = \frac{1}{\text{tr}[L]}L, \quad \rho^I = \frac{1}{\text{tr}[I]}I, \quad \rho^F = \frac{1}{\text{tr}[F]}F. \tag{H12}$$

Both ρ^L and ρ^I are diagonal in the standard basis. In consequence, the output fidelity on these states is upper bounded by 0.5. Hence by Lemma 2 we see that

$$f_{\text{opt}}(\rho_{AB}(p,p_d)) \leqslant f_{\text{opt}}(\rho^F). \tag{H13}$$

Note that ρ^F only has support on a bipartite two-qubit subspace:

$$\rho^F = \tfrac{1}{2}(|01\rangle\langle 01|_A \otimes |10\rangle\langle 10|_B + d|01\rangle\langle 10|_A \otimes |10\rangle\langle 01|_B + d|10\rangle\langle 01|_A \otimes |01\rangle\langle 10|_B + |10\rangle\langle 10|_A \otimes |01\rangle\langle 01|_B). \tag{H14}$$

Hence, Alice and Bob can redefine their state according to

$$|01\rangle_A \to |0\rangle_A, \quad |10\rangle_A \to |1\rangle_A, \quad |01\rangle_B \to |1\rangle_B, \quad |10\rangle_B \to |0\rangle_B. \tag{H15}$$

Under such local relabeling the state ρ^F becomes

$$\rho^F = p_d|\Phi^+\rangle\langle\Phi^+| + (1-p_d)|\Phi^-\rangle\langle\Phi^-|. \tag{H16}$$

We know from [47] that it is not possible to increase the fidelity of the state in Eq. (H16) through local filtering. In consequence,

$$f_{\rm opt}(\rho_{AB}(p,p_d)) \leqslant p_d. \tag{H17}$$

Since the output fidelity of EPL-D is exactly p_d, EPL-D achieves the optimal fidelity. Now we show that it achieves this output fidelity with the highest possible probability of success. From Lemma 3, it follows that this probability of success is upper bounded by the relative weight of ρ_F in $\rho_{AB}(p,p_d)$, which is $p^2/2$. Since EPL-D achieves the output fidelity of p_d with success probability $p^2/2$, we can conclude that it is also optimal with respect to probability of success. Hence EPL-D is fidelity-optimal for the EPL remote entanglement generation.

2. EPL-D is distillation-optimal

Let us consider the distillable entanglement of the state in Eq. (H3). Unfortunately, there is no straightforward way of calculating distillable entanglement. However, distillable entanglement is upper bounded by the relative entropy of entanglement [55]:

$$E_R(\rho) = \min_{\sigma\in{\rm SEP}} S(\rho\|\sigma), \tag{H18}$$

where $S(\rho\|\sigma)$ is the relative entropy defined as

$$S(\rho\|\sigma) = {\rm tr}[\rho\log\rho] - {\rm tr}[\rho\log\sigma]. \tag{H19}$$

Moreover, $S(\rho\|\sigma)$ for any $\sigma\in$ SEP is an upper bound on $E_R(\rho)$ and, in consequence, on $E_D(\rho)$. Consider the separable state

$$\sigma_{AB}^{\rm SEP}(p) = \frac{p^2}{4}P_{{\rm odd}_{A1B1}}\otimes P_{{\rm odd}_{A2B2}} + \frac{(1-p)p}{2}[|11\rangle\langle 11|_{A1B1}\otimes P_{{\rm odd}_{A2B2}} \tag{H20}$$

$$+ P_{{\rm odd}_{A1B1}}\otimes|11\rangle\langle 11|_{A2B2}] + (1-p)^2|11\rangle\langle 11|_{A1B1}\otimes|11\rangle\langle 11|_{A2B2}. \tag{H21}$$

Then we can calculate

$$S\big(\rho_{AB}(p,p_d)\|\sigma_{AB}^{\rm SEP}(p)\big) = \frac{p^2}{2}[1-h(p_d)], \tag{H22}$$

where h denotes the binary entropy function. We can conclude that $E_D(\rho_{AB}(p,p_d)) \leqslant \frac{p^2}{2}[1-h(p_d)]$.

Now, we note that a possible distillation scheme would be to first perform the EPL-D protocol on the individual copies of the state in Eq. (H3) and then perform the optimal achievable distillation procedure on the output states. Hence it is possible to distill EPR states from the states in Eq. (H3) at a rate given by

$$R = p_{\rm succ,EPL\text{-}D}E_D(\eta_{\hat{A}\hat{B}}(p_d)). \tag{H23}$$

The success probability of EPL-D is $\frac{p^2}{2}$ and the distillable entanglement of rank-two Bell diagonal states is [48]

$$E_D(\eta_{\hat{A}\hat{B}}(p_d)) = 1 - h(p_d). \tag{H24}$$

Hence we can conclude that $E_D(\rho_{AB}(p,p_d)) = \frac{p^2}{2}[1-h(p_d)]$ and so $E_D(\rho_{AB}(p,p_d)) = p_{\rm succ,EPL\text{-}D}E_D(\eta_{\hat{A}\hat{B}}(p_d))$. This proves that EPL-D is distillation-optimal for EPL remote entanglement generation scheme.

[1] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller, Quantum Repeaters: The Role of Imperfect Local Operations in Quantum Communication, Phys. Rev. Lett. **81**, 5932 (1998).

[2] S. Bratzik, S. Abruzzo, H. Kampermann, and D. Bruß, Quantum repeaters and quantum key distribution: The impact of entanglement distillation on the secret key rate, Phys. Rev. A **87**, 062335 (2013).

[3] S. Guha, H. Krovi, C. A. Fuchs, Z. Dutton, J. A. Slater, C. Simon, and W. Tittel, Rate-loss analysis of an efficient quantum repeater architecture, Phys. Rev. A **92**, 022357 (2015).

[4] K. G. H. Vollbrecht, C. A. Muschik, and J. I. Cirac, Entanglement Distillation by Dissipation and Continuous Quantum Repeaters, Phys. Rev. Lett. **107**, 120502 (2011).

[5] W. J. Munro, K. Azuma, K. Tamaki, and K. Nemoto, Inside quantum repeaters, IEEE J. Sel. Top. Quantum Electron. **21**, 78 (2015).

[6] N. H. Nickerson, Y. Li, and S. C. Benjamin, Topological quantum computing with a very noisy network and local error rates approaching one percent, Nat. Commun. **4**, 1756 (2013).

[7] N. H. Nickerson, J. F. Fitzsimons, and S. C. Benjamin, Freely Scalable Quantum Technologies Using Cells of 5-to-50 Qubits with Very Lossy and Noisy Photonic Links, Phys. Rev. X **4**, 041041 (2014).

[8] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, Purification of Noisy Entanglement

and Faithful Teleportation via Noisy Channels, Phys. Rev. Lett. **76**, 722 (1996).

[9] D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera, Quantum Privacy Amplification and the Security of Quantum Cryptography Over Noisy Channels, Phys. Rev. Lett. **77**, 2818 (1996).

[10] Z. Zhao, J.-W. Pan, and M. Zhan, Practical scheme for entanglement concentration, Phys. Rev. A **64**, 014301 (2001).

[11] T. Yamamoto, M. Koashi, and N. Imoto, Concentration and purification scheme for two partially entangled photon pairs, Phys. Rev. A **64**, 012304 (2001).

[12] J.-W. Pan, C. Simon, Č. Brukner, and A. Zeilinger, Entanglement purification for quantum communication, Nature (London) **410**, 1067 (2001).

[13] E. T. Campbell and S. C. Benjamin, Measurement-Based Entanglement Under Conditions of Extreme Photon Loss, Phys. Rev. Lett. **101**, 130502 (2008).

[14] P. G. Kwiat, S. Barraza-Lopez, A. Stefanov, and N. Gisin, Experimental entanglement distillation and hidden non-locality, Nature (London) **409**, 1014 (2001).

[15] Z. Zhao, T. Yang, Y.-A. Chen, A.-N. Zhang, and J.-W. Pan, Experimental Realization of Entanglement Concentration and a Quantum Repeater, Phys. Rev. Lett. **90**, 207901 (2003).

[16] R. Reichle, D. Leibfried, E. Knill, J. Britton, R. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Experimental purification of two-atom entanglement, Nature (London) **443**, 838 (2006).

[17] H. Takahashi, J. S. Neergaard-Nielsen, M. Takeuchi, M. Takeoka, K. Hayasaka, A. Furusawa, and M. Sasaki, Entanglement distillation from Gaussian input states, Nat. Photonics **4**, 178 (2010).

[18] N. Kalb, A. A. Reiserer, P. C. Humphreys, J. J. Bakermans, S. J. Kamerling, N. H. Nickerson, S. C. Benjamin, D. J. Twitchen, M. Markham, and R. Hanson, Entanglement distillation between solid-state quantum network nodes, Science **356**, 928 (2017).

[19] W. Dür and H. J. Briegel, Entanglement purification and quantum error correction, Rep. Prog. Phys. **70**, 1381 (2007).

[20] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Mixed-state entanglement and quantum error correction, Phys. Rev. A **54**, 3824 (1996).

[21] E. M. Rains, A semidefinite program for distillable entanglement, IEEE Trans. Inf. Theory **47**, 2921 (2001).

[22] G. Vidal and R. F. Werner, Computable measure of entanglement, Phys. Rev. A **65**, 032314 (2002).

[23] M. B. Plenio, Logarithmic Negativity: A Full Entanglement Monotone That Is Not Convex, Phys. Rev. Lett. **95**, 090503 (2005).

[24] X. Wang and R. Duan, Improved semidefinite programming upper bound on distillable entanglement, Phys. Rev. A **94**, 050301 (2016).

[25] M. Tomamichel, M. Berta, and J. M. Renes, Quantum coding with finite resources, Nat. Commun. **7**, 11419 (2016).

[26] F. Buscemi and N. Datta, Distilling entanglement from arbitrary resources, J. Math. Phys. **51**, 102201 (2010).

[27] K. Fang, X. Wang, M. Tomamichel, and R. Duan, Non-asymptotic entanglement distillation, arXiv:1706.06221.

[28] F. G. Brandao and N. Datta, One-shot rates for entanglement manipulation under non-entangling maps, IEEE Trans. Inf. Theory **57**, 1754 (2011).

[29] L.-M. Duan, M. Lukin, J. I. Cirac, and P. Zoller, Long-distance quantum communication with atomic ensembles and linear optics, Nature (London) **414**, 413 (2001).

[30] S. D. Barrett and P. Kok, Efficient high-fidelity quantum computation using matter qubits and linear optics, Phys. Rev. A **71**, 060310 (2005).

[31] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, A complete family of separability criteria, Phys. Rev. A **69**, 022308 (2004).

[32] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, Distinguishing Separable and Entangled States, Phys. Rev. Lett. **88**, 187904 (2002).

[33] Available online: https://github.com/StephanieWehner/EntanglementDist.jl.

[34] N. Gisin, Hidden quantum nonlocality revealed by local filters, Phys. Lett. A **210**, 151 (1996).

[35] E. M. Rains, Bound on distillable entanglement, Phys. Rev. A **60**, 179 (1999).

[36] E. M. Rains, Rigorous treatment of distillable entanglement, Phys. Rev. A **60**, 173 (1999).

[37] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004).

[38] M. Navascues, M. Owari, and M. B. Plenio, Power of symmetric extensions for entanglement detection, Phys. Rev. A **80**, 052306 (2009).

[39] A. C. Doherty, Entanglement and the shareability of quantum states, J. Phys. A **47**, 424004 (2014).

[40] M. M. Wolf, Quantum channels and operations: Guided tour. Available online: https://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf.

[41] J. Dehaene, M. Van den Nest, B. De Moor, and F. Verstraete, Local permutations of products of Bell states and entanglement distillation, Phys. Rev. A **67**, 022310 (2003).

[42] L. Ruan, W. Dai, and M. Z. Win, Adaptive recurrence quantum entanglement distillation for two-Kraus-operator channels, Phys. Rev. A **97**, 052332 (2018).

[43] E. T. Campbell, How to exploit local information when distilling entanglement, Int. J. Quantum Inf. **8**, 161 (2010).

[44] C. Cabrillo, J. Cirac, P. Garcia-Fernandez, and P. Zoller, Creation of entangled states of distant atoms by interference, Phys. Rev. A **59**, 1025 (1999).

[45] A. Kent, N. Linden, and S. Massar, Optimal Entanglement Enhancement for Mixed States, Phys. Rev. Lett. **83**, 2656 (1999).

[46] F. Verstraete, J. Dehaene, and B. DeMoor, Local filtering operations on two qubits, Phys. Rev. A **64**, 010101 (2001).

[47] F. Verstraete and H. Verschelde, Optimal Teleportation with a Mixed State of Two Qubits, Phys. Rev. Lett. **90**, 097901 (2003).

[48] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi, Fundamental limits of repeaterless quantum communications, Nat. Commun. **8**, 15043 (2017).

[49] M. Horodecki, P. Horodecki, and R. Horodecki, General teleportation channel, singlet fraction, and quasidistillation, Phys. Rev. A **60**, 1888 (1999).

[50] W. Pfaff, B. Hensen, H. Bernien, S. B. van Dam, M. S. Blok, T. H. Taminiau, M. J. Tiggelman, R. N. Schouten, M. Markham, D. J. Twitchen *et al.*, Unconditional quantum teleportation between distant solid-state quantum bits, Science **345**, 532 (2014).

[51] Y. Shi and X. Wu, Epsilon-net method for optimizations over separable states, Theoretical Computer Science **598**, 51 (2015).

[52] J. Datorro, *Convex Optimization Euclidean Distance Geometry* (Meboo Publishing, Palo Alto, 2015).

[53] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, J. Opt. B **7**, S347 (2005).

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

[55] M. Horodecki, P. Horodecki, and R. Horodecki, Limits for Entanglement Measures, Phys. Rev. Lett. **84**, 2014 (2000).

[56] M. Horodecki, P. Horodecki, and R. Horodecki, Inseparable Two Spin-$1/2$ Density Matrices Can Be Distilled to a Singlet Form, Phys. Rev. Lett. **78**, 574 (1997).

[57] It must be noted that there also exists a general procedure for distilling any inseparable two-qubit state, and in particular any two-qubit state whose fidelity to any maximally entangled state is smaller than or equal to half and which therefore cannot be distilled using DEJMPS or BBPSSW; see [47,56].