Adaptive recurrence quantum entanglement distillation for two-Kraus-operator channels

Liangzhong Ruan, Wenhan Dai, and Moe Z. Win
Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 32-D608, Cambridge, Massachusetts 02139, USA

(Received 9 December 2016; published 29 May 2018)

Quantum entanglement serves as a valuable resource for many important quantum operations. A pair of entangled qubits can be shared between two agents by first preparing a maximally entangled qubit pair at one agent, and then sending one of the qubits to the other agent through a quantum channel. In this process, the deterioration of entanglement is inevitable since the noise inherent in the channel contaminates the qubit. To address this challenge, various quantum entanglement distillation (QED) algorithms have been developed. Among them, recurrence algorithms have advantages in terms of implementability and robustness. However, the efficiency of recurrence QED algorithms has not been investigated thoroughly in the literature. This paper puts forth two recurrence QED algorithms that adapt to the quantum channel to tackle the efficiency issue. The proposed algorithms have guaranteed convergence for quantum channels with two Kraus operators, which include phase-damping and amplitude-damping channels. Analytical results show that the convergence speed of these algorithms is improved from linear to quadratic and one of the algorithms achieves the optimal speed. Numerical results confirm that the proposed algorithms significantly improve the efficiency of QED.

DOI: 10.1103/PhysRevA.97.052332

I. INTRODUCTION

Quantum entanglement shared by remote agents serves as a valuable resource for many important applications of quantum computation and quantum information [1–3], such as secret key distribution [4–6], dense coding [7–9], and teleportation [10–12]. With the assistance of entanglement, the capacity of quantum channels can be increased, particularly when the channels are very noisy [13–17]. Entanglement also enables quantum relay, and therefore is a keystone of long-distance quantum communication [18]. To establish entanglement between two remote agents, one agent can locally generate a maximally entangled qubit pair and send one of the qubits to the other agent through a quantum channel. However, the noise inherent in the channel will contaminate the qubit during the transmission, thereby deteriorating the entanglement. To address this problem, quantum entanglement distillation (QED) algorithms [19–29] have been proposed to generate highly entangled shared qubit pairs from many contaminated ones via local operations and classical communication (LOCC). Since high-quality entanglement is the keystone in many important applications of quantum computation and quantum information, QED has become an essential building block for the development of quantum networks [30–32].

In the pioneering work [19], two influential QED algorithms were proposed and are now known as the recurrence algorithm and the asymptotic algorithm. Recurrence algorithms [20–22] operate separately on every two qubit pairs, improving the quality of entanglement in one pair at the expense of the other pair, which is then discarded. The algorithms keep repeating this operation to progressively improve the quality of entanglement in the kept qubit pairs. Asymptotic algorithms [23–25] operate on a large number of qubit pairs, detecting ones that are not in the targeted state by measuring a subset of the qubit pairs, and then transforming those in the undesired state to the targeted state. Later, it was recognized that there is a duality between QED and quantum error correction (QEC) [26] when one-way classical communication is involved. The connection between QED and QEC was further explored in scenarios involving two-way classical communication, enabling code-based QED algorithms [27–29]. These algorithms operate on a few qubit pairs, look for the error syndrome using measurements specified by the error correction code, and then correct the errors to restore entanglement.

Asymptotic algorithms have revealed important theoretical insights, but these algorithms require agents that have the capability of processing a large number of qubits. Code-based algorithms require agents to have the capability of processing only a few qubits, but the number of errors that can be corrected is limited by the Hamming distance of the codewords. Designing QEC codes with large Hamming distance is challenging since the creation of information redundancy, the main mechanism adopted in classical error correction codes, is not possible in quantum codes due to the no-cloning theorem [33–36]. Hence, these algorithms do not apply to scenarios with strong noise in the channel. Recurrence algorithms require agents to have the capability of processing only a few qubits and can generate maximally entangled qubit pairs even in strong noise scenarios. This is because the recurrence algorithms can mitigate stronger noise by performing more rounds of distillations. In fact, the recurrence algorithm proposed in [19] can distill contaminated qubit pairs into maximally entangled qubit pairs as long as the initial fidelity of the contaminated qubit pairs with respect to the targeted state is greater than 0.5.

Building large-scale quantum circuits operating on many qubits is challenging [37–39]; even the error rates of two-qubit operations are significantly higher than those of one-qubit operations [40]. In this perspective, recurrence algorithms are favorable for implementation as they require operations only

on a few (typically one or two) qubits and are robust to strong noise in the quantum channel. On the other hand, since at least half of the entangled qubit pairs are discarded in each round of distillation, the efficiency of the recurrence algorithms decreases dramatically with the number of rounds.[1] To address this challenge, the quantum privacy amplification (QPA) algorithm was proposed in [20], and was shown numerically to require fewer rounds of distillation than the algorithm in [19] for contaminated qubit pairs with a specific set of initial states. However, the performance of QPA algorithm was not characterized analytically. In fact, another set of initial states was found in [21] for which the QPA algorithm was less efficient than the algorithm in [19]. In [21], the design of distillation operations was formulated into an optimization problem, which was inherently nonconvex, and consequently the optimal solution was not found.

We envision that a key enabler to designing efficient recurrence QED algorithms is to make them adaptive to quantum channels. Intuitively, compared to general algorithms, QED algorithms that adapt to channel-specific noise will better mitigate such noise and hence distill more efficiently. In fact, it has been observed that knowing the channel benefits the performance of quantum error recovery [41], and channel-adaptive QEC schemes that outperform prior ones [42,43] have been designed.

In this paper, we focus on two-Kraus-operator (TKO) channels, a class that covers several typical quantum channels, e.g., phase-damping and amplitude-damping channels. The phase-damping channels describe the decoherence process of a photon traveling through a waveguide, and the amplitude-damping channels model the decay of an excited atom due to spontaneous emission (see [44], Sec. 3.4] and [45], Sec.8.3]). To achieve efficient distillation, we develop two adaptive recurrence QED algorithms, which adapt to the channel by employing a remote shared-state preparation (RSSP) method.[2] The contributions of this paper are (1) characterization of the structure of TKO channels; (2) characterization of the optimal fidelity that can be achieved by performing LOCC on two qubit pairs affected by TKO channels; and (3) design of adaptive recurrence QED algorithms, which improve the convergence speed of fidelity from linear to quadratic, and one of them achieves the optimal speed.

Notations

a, $\boldsymbol{a}$, and $\boldsymbol{A}$ represent scalar, vector, and matrices, respectively. pha$\{\cdot\}$ denotes the phase of a complex number. $(\cdot)^\dagger$, rank$\{\cdot\}$, det$\{\cdot\}$, and tr$\{\cdot\}$ denote the Hermitian transpose, rank, determinant, and trace of a matrix, respectively. $\mathrm{tr}_{i,j}\{\cdot\}$ denotes the partial trace with respect to to the ith and jth qubits in the system. span$(\cdot)$ denotes the linear space spanned by a set of vectors. $\mathbb{I}_n$ denotes the $n \times n$ identity matrix, and $\imath$ is the unit imaginary number.

II. SYSTEM MODEL

Consider two remote agents, Alice and Bob, connected by a quantum channel and a two-way classical channel. When Alice transmits a qubit with density matrix $\boldsymbol{\rho}_0$, the density matrix of the qubit received by Bob is given by

$$\boldsymbol{\rho} = \sum_{k=1}^{K} \boldsymbol{C}_k \boldsymbol{\rho}_0 \boldsymbol{C}_k^\dagger, \tag{1}$$

where the Kraus operators $\{\boldsymbol{C}_k\}$ representing the noisy quantum channel satisfy

$$\sum_{k=1}^{K} \boldsymbol{C}_k^\dagger \boldsymbol{C}_k = \mathbb{I}_2. \tag{2}$$

Since a qubit is a two-dimensional system, the number of the Kraus operators $K \leqslant 2^2 = 4$.[3] When $K = 1$, the channel is noiseless, and hence QED is not needed. For the class of TKO channels, $K = 2$.

Suppose Alice and Bob wish to obtain maximally entangled qubit pairs with density matrix $\boldsymbol{\rho}_0 = |\Phi^+\rangle\langle\Phi^+|$, where $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. To achieve this task, Alice locally prepares qubit pairs, each with density matrix $\boldsymbol{\rho}_0$, then sends the second qubit in each pair through the noisy channel. Then the density matrix of the two remote qubits becomes

$$\boldsymbol{\rho} = \sum_{k=1}^{2} (\mathbb{I}_2 \otimes \boldsymbol{C}_k)\, \boldsymbol{\rho}_0\, (\mathbb{I}_2 \otimes \boldsymbol{C}_k)^\dagger. \tag{3}$$

Alice and Bob then adopt a recurrence QED algorithm outlined in Fig. 1. In each round of distillation, the agents separately operate on every two qubit pairs kept in the previous round, perform LOCC, and attempt to improve the quality of entanglement in one of the qubit pairs at the expense of the other pair, which is then discarded (agents may discard both pairs when this operation is unsuccessful). The objective of the algorithm is to generate qubit pairs with density matrix $\boldsymbol{\rho}^*$ close to the targeted state, i.e.,

$$\langle\Phi^+|\, \boldsymbol{\rho}\, |\Phi^+\rangle \approx 1$$

where $\langle\Phi^+|\, \boldsymbol{\rho}\, |\Phi^+\rangle$ is the fidelity of a density matrix $\boldsymbol{\rho}$ and the targeted state $|\Phi^+\rangle$.

III. DESIGN OF THE ADAPTIVE RECURRENCE QED ALGORITHM

A. Characterization of TKO channels

Prior to designing adaptive QED algorithms, it is crucial to understand the effect of noisy quantum channels on the

[1] The efficiency of QED algorithms is measured in terms of *yield*, which is defined as the ratio between the number of maximally entangled output qubit pairs and the number of contaminated input qubit pairs.

[2] This method is akin to remote state preparation methods [46,47], in which two remote agents employ LOCC to prepare a quantum state at one of the agents. The proposed RSSP method uses LOCC to prepare a state shared by both agents.

[3] Quantum operators on n-dimensional systems are $n \times n$ matrices, and hence lie in an n^2-dimensional space. Thus, from [44], Sec 3.3], if a channel for such systems is represented with more than n^2 operators, there always exists an equivalent representation with no more than n^2 nonzero operators.

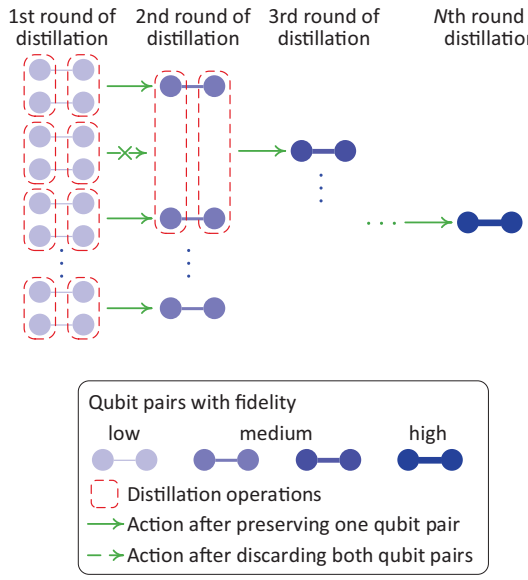

FIG. 1. The structure of recurrence QED algorithms.

entanglement between qubits. This can be accomplished by determining the structure of the noisy quantum channels. In particular, the structure of TKO channels is provided by the following lemma.

Lemma 1. Structure of TKO channels. For every single-qubit TKO channel, there exist unitary matrices $\boldsymbol{U}$ and $\boldsymbol{V} \in \mathbb{C}^{2\times 2}$ and scalars $p \in [0,1], \zeta \in [0,1]$, and $\eta \in \mathbb{C}$ with $|\eta|^2 + \zeta^2 = 1$ such that the channel can be represented by

$$\boldsymbol{C}_1 = \boldsymbol{U}\begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{bmatrix}\boldsymbol{V}^\dagger, \quad \boldsymbol{C}_2 = \boldsymbol{U}\begin{bmatrix} 0 & \eta\sqrt{p} \\ 0 & \zeta\sqrt{p} \end{bmatrix}\boldsymbol{V}^\dagger. \tag{4}$$

Proof. The proof is given in Appendix A. □

Local unitary operations $\boldsymbol{U}$ and $\boldsymbol{V}$ do not affect the amount of entanglement [48]. In particular, for any $\boldsymbol{U}$ and $\boldsymbol{V}$, Bob and Alice can, respectively, perform local unitary operations $\boldsymbol{U}^\dagger$ and $\boldsymbol{V}$ and obtain an equivalent channel with $\tilde{\boldsymbol{U}} = \boldsymbol{U}^\dagger\boldsymbol{U} = \mathbb{I}_2$ and $\tilde{\boldsymbol{V}} = \boldsymbol{V}^\dagger\boldsymbol{V} = \mathbb{I}_2$. Therefore, without loss of generality, $\boldsymbol{U}$ and $\boldsymbol{V}$ are assumed to be $\mathbb{I}_2$ in the following analysis. In this case, the channel can be represented by

$$\boldsymbol{C}_1 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{bmatrix}, \quad \boldsymbol{C}_2 = \begin{bmatrix} 0 & \eta\sqrt{p} \\ 0 & \zeta\sqrt{p} \end{bmatrix}. \tag{5}$$

Remark 1. The effect of the TKO channel on entanglement. From (5), $\mathrm{rank}\{\boldsymbol{C}_2\} = 1$. Hence, $\boldsymbol{C}_2$ can be written as in (A1). In this case, for any qubit pairs with density matrix $\boldsymbol{\sigma}_0$, conditional on that $\boldsymbol{C}_2$ operates, the density matrix of the qubit pair after the second qubit goes through the channel becomes

$$\boldsymbol{\sigma} = \frac{(\mathbb{I}_2 \otimes \boldsymbol{C}_2)\,\boldsymbol{\sigma}_0\,(\mathbb{I}_2 \otimes \boldsymbol{C}_2)^\dagger}{\mathrm{tr}\{(\mathbb{I}_2 \otimes \boldsymbol{C}_2)\,\boldsymbol{\sigma}_0\,(\mathbb{I}_2 \otimes \boldsymbol{C}_2)^\dagger\}} = \boldsymbol{\sigma}_\mathrm{A} \otimes \boldsymbol{\sigma}_\mathrm{B},$$

where $\boldsymbol{\sigma}_\mathrm{B} = |i\rangle\langle i|$, with $|i\rangle = \eta|0\rangle + \zeta|1\rangle$. Therefore, when $\boldsymbol{C}_2$ operates, $\boldsymbol{\sigma}$ is a separable state, which implies that all entanglement between the two qubits is destroyed. The behavior of $\boldsymbol{C}_2$ is mainly characterized by parameters p and η.

(1) The role of p. $p \in [0,1]$ is the strength of $\boldsymbol{C}_2$, which is proportional to the probability that $\boldsymbol{C}_2$ operates on a qubit. The parameter p can be thought of as the severity of noise in the channel since it characterizes the extent that the channel deteriorates the entanglement. The larger the p, the more the entanglement is destroyed. In particular, all entanglement is preserved when $p = 0$, and the opposite is true when $p = 1$.

(2) The role of η. $\arcsin(|\eta|) \in [0, \frac{\pi}{2}]$ is the angle between the image, i.e., $\mathrm{span}(\eta|0\rangle + \zeta|1\rangle)$, and the coimage, i.e., $\mathrm{span}(|1\rangle)$, of $\boldsymbol{C}_2$. This characterizes the angle at which $\boldsymbol{C}_2$ rotates the state of a qubit, and hence indicates the type of the channel. In particular, the channel is phase damping when $\arcsin(|\eta|) = 0$, i.e., $\eta = 0$, and amplitude damping when $\arcsin(|\eta|) = \frac{\pi}{2}$, i.e., $|\eta| = 1$.

The channel properties described above are summarized in Fig. 2. □

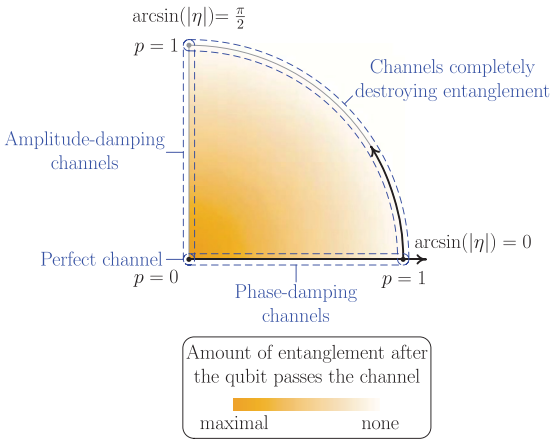

FIG. 2. Characterization of TKO channels.

All entanglement is destroyed after the second qubit passes through a TKO channel with $p = 1$. Therefore, the interesting case for QED is $p \in [0,1)$. The following theorem characterizes the structure of the density matrix $\boldsymbol{\rho}$ in this case.

Theorem 1. Structure of $\boldsymbol{\rho}$. Consider the density matrix of a qubit pair after the second qubit passes through a TKO channel represented by (5). When $p < 1$, there exist local unitary operators $\boldsymbol{U}_\mathrm{A}$ and $\boldsymbol{U}_\mathrm{B}$ such that

$$\breve{\boldsymbol{\rho}} = (\boldsymbol{U}_\mathrm{A} \otimes \boldsymbol{U}_\mathrm{B})\,\boldsymbol{\rho}\,(\boldsymbol{U}_\mathrm{A} \otimes \boldsymbol{U}_\mathrm{B})^\dagger = F|\mu\rangle\langle\mu| + (1-F)|\nu\rangle\langle\nu|, \tag{6}$$

where

$$|\mu\rangle = \alpha|00\rangle + \beta|11\rangle, \tag{7}$$

$$|\nu\rangle = \gamma|01\rangle + \delta e^{\imath\theta}|10\rangle \tag{8}$$

with θ a certain constant in $[0, 2\pi)$, and

$$F = \frac{1}{2} + \frac{1}{2}\sqrt{(1-p)(1-|\eta|^2 p)}, \tag{9}$$

$$\alpha = \sqrt{\frac{1}{2} + \frac{|\eta|p}{4F}}, \quad \beta = \sqrt{\frac{1}{2} - \frac{|\eta|p}{4F}}, \tag{10}$$

$$\gamma = \sqrt{\frac{1}{2} - \frac{|\eta|p}{4(1-F)}}, \quad \delta = \sqrt{\frac{1}{2} + \frac{|\eta|p}{4(1-F)}}. \tag{11}$$

Proof. The proof is given in Appendix B. □

Remark 2. The role of $\boldsymbol{U}_\mathrm{A}$ and $\boldsymbol{U}_\mathrm{B}$. Equations (6)–(8) show that by applying the properly designed local unitary operators

$\boldsymbol{U}_{\mathrm{A}}$ and $\boldsymbol{U}_{\mathrm{B}}$, the density matrix $\boldsymbol{\rho}$ can be transformed into $\breve{\boldsymbol{\rho}}$ with all its eigenvectors written in the computational basis, i.e., $\{|0\rangle,|1\rangle\}$ via the Schmidt decomposition. This transformation enables the simplification for both analysis and algorithm design in the following sections. In particular, $\boldsymbol{U}_{\mathrm{A}}$ and $\boldsymbol{U}_{\mathrm{B}}$ will be employed by Alice and Bob, respectively, on their individual qubits before the recurrent distillation operations begin, and hence will be referred to as the predistillation unitary operators. □

B. Characterization of the optimal fidelity

This section proves the optimal fidelity that can be achieved by performing appropriate LOCC. Consider recurrence QED algorithms outlined in Fig. 1, which (1) perform LOCC on two qubit pairs with density matrix $\breve{\boldsymbol{\rho}}$ given by equations (6) and (2) keep at most one pair. Since operations performed by agents are local, they can be expressed as $\boldsymbol{N}_{\mathrm{A}}^{(k)} \otimes \boldsymbol{N}_{\mathrm{B}}^{(k)}$, $k \in \{1,2,\ldots,K\}$, satisfying $\sum_{k=1}^{K} (\boldsymbol{N}_{\mathrm{A}}^{(k)})^{\dagger} \boldsymbol{N}_{\mathrm{A}}^{(k)} = \sum_{k=1}^{K} (\boldsymbol{N}_{\mathrm{B}}^{(k)})^{\dagger} \boldsymbol{N}_{\mathrm{B}}^{(k)} = \mathbb{I}_4$. Without loss of generality, assume that agents keep the first qubit pair conditioned on the event that one of the first $\tilde{K}$ operators acts on the four qubits. Then after the LOCC, the density matrix of the first qubit pair is given by

$$\breve{\boldsymbol{\rho}} = \frac{\mathrm{tr}_{2,4}\left\{\sum_{k=1}^{\tilde{K}} \left(\boldsymbol{N}_{\mathrm{A}}^{(k)} \otimes \boldsymbol{N}_{\mathrm{B}}^{(k)}\right)\left(\boldsymbol{P}\breve{\boldsymbol{\rho}} \otimes \breve{\boldsymbol{\rho}}\, \boldsymbol{P}^{\dagger}\right)\left(\boldsymbol{N}_{\mathrm{A}}^{(k)} \otimes \boldsymbol{N}_{\mathrm{B}}^{(k)}\right)^{\dagger}\right\}}{\mathrm{tr}\left\{\sum_{k=1}^{\tilde{K}} \left(\boldsymbol{N}_{\mathrm{A}}^{(k)} \otimes \boldsymbol{N}_{\mathrm{B}}^{(k)}\right)\left(\boldsymbol{P}\breve{\boldsymbol{\rho}} \otimes \breve{\boldsymbol{\rho}}\, \boldsymbol{P}^{\dagger}\right)\left(\boldsymbol{N}_{\mathrm{A}}^{(k)} \otimes \boldsymbol{N}_{\mathrm{B}}^{(k)}\right)^{\dagger}\right\}},$$

where

$$\boldsymbol{P} = \mathbb{I}_2 \otimes (|00\rangle\langle 00| + |10\rangle\langle 01| + |01\rangle\langle 10| + |11\rangle\langle 11|) \otimes \mathbb{I}_2$$

is the permutation operator that switches the second and third qubits. With this operator, the joint density matrix $\boldsymbol{\rho}_{\mathrm{J}} = \boldsymbol{P}\breve{\boldsymbol{\rho}} \otimes \breve{\boldsymbol{\rho}}\boldsymbol{P}^{\dagger}$ corresponds to four qubits, where the first two belong to Alice and last two belong to Bob.

Denote F^* as the optimal fidelity that can be achieved with initial density matrix $\breve{\boldsymbol{\rho}}$ via all possible LOCC, i.e.,

$$F^* = \max_{\{\boldsymbol{N}_{\mathrm{A}}^{(k)}, \boldsymbol{N}_{\mathrm{B}}^{(k)}\}_{k=1}^{\tilde{K}} \in \mathcal{F}} \langle \Phi^+ | \breve{\boldsymbol{\rho}} | \Phi^+ \rangle, \tag{12}$$

where $\mathcal{F}$ denotes the set of all possible LOCC.

The characterization of the optimal fidelity F^* is challenging because it involves general TKO channels and arbitrary LOCC. These two issues are tackled by the following lemmas. Lemma 2 characterizes the relationship between the F^* for general TKO channels and that for the special case of phase-damping channels. Lemma 3 exploits the property of separable operators to determine the set of attainable density matrices of the first qubit pair.

Lemma 2. Simplification to phase damping. Express the optimal fidelity F^* explicitly as a function of the density matrix parameters in (9)–(11), i.e.,

$$F^* = f(F,\alpha,\beta,\gamma,\delta,\theta).$$

If the optimal fidelity for phase-damping channels is upper bounded by

$$f\left(F, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right) \leqslant \frac{F^2}{F^2 + (1-F)^2}, \quad \forall F \in \left(\frac{1}{2}, 1\right]$$

then the optimal fidelity for generic TKO channels satisfies

$$f(F,\alpha,\beta,\gamma,\delta,\theta) \leqslant \frac{F^2}{F^2 + (1-F)^2\left(\frac{\gamma\delta}{\alpha\beta}\right)^2}, \tag{13}$$

$\forall F$, α, β, γ, δ, and θ satisfying (9)–(11).

Proof. The proof is given in Appendix C. □

Lemma 3. Density matrix after arbitrary separable operation. For phase-damping channels, after an arbitrary separable operator acts on two qubit pairs, the density matrix of the kept qubit pair can be expressed as

$$\breve{\boldsymbol{\rho}} = \frac{\sum_{i=1}^{4} C_i \boldsymbol{\psi}^{(i)} \boldsymbol{\psi}^{(i)\dagger}}{\sum_{i=1}^{4} C_i \boldsymbol{\psi}^{(i)\dagger} \boldsymbol{\psi}^{(i)}}, \tag{14}$$

where $C_1 = F^2$, $C_2 = C_3 = F(1-F)$, $C_4 = (1-F)^2$,

$$\boldsymbol{\psi}^{(i)} = \begin{bmatrix} w_{11} & x_{11} & y_{11} & z_{11} \\ w_{12} & x_{12} & y_{12} & z_{12} \\ w_{21} & x_{21} & y_{21} & z_{21} \\ w_{22} & x_{22} & y_{22} & z_{22} \end{bmatrix} \boldsymbol{v}^{(i)}, \tag{15}$$

in which s_{ij}, $s \in \{w,x,y,z\}$, and $i,j \in \{1,2\}$ are complex numbers satisfying

$$s_{11}s_{22} = s_{12}s_{21}, \tag{16}$$

$\boldsymbol{v}^{(i)}$ is the ith column of the unitary matrix

$$\boldsymbol{V} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}, \tag{17}$$

and $\sum_{i=1}^{4} C_i \boldsymbol{\psi}^{(i)\dagger} \boldsymbol{\psi}^{(i)} > 0$.

Proof. The proof is given in Appendix D. □

With the issues of general TKO channels and arbitrary LOCC addressed, the optimal fidelity F^* can now be characterized.

Theorem 2. Optimal fidelity. Consider the density matrix $\breve{\boldsymbol{\rho}}$, given in (6), of a pair of entangled qubits shared by agents via a TKO channel. Then the optimal fidelity of the kept qubit pair after performing LOCC is given by

$$F^* = \frac{F^2}{F^2 + (1-F)^2\left(\frac{\gamma\delta}{\alpha\beta}\right)^2}. \tag{18}$$

Proof. The proof is given in Appendix E. □

Remark 3. Key channel parameters and the optimal fidelity. Equation (18) describes the optimal fidelity as a function of the parameters of the density matrix. To understand how parameters of the channel affect the optimal fidelity F^*, one can substitute (9)–(11) into (18) to obtain

$$F^* = \frac{1}{2} + \frac{\sqrt{(1-p)(1-|\eta|^2 p)}}{(1-p) + (1-|\eta|^2 p)}. \tag{19}$$

By taking the derivative of (19) with respect to p and $|\eta|$, respectively, it can be verified that F^* is a decreasing function of p and an increasing function of $|\eta|$. An intuitive understanding of such trends can be obtained by recalling Remark 1 and Fig. 2. Operator $\boldsymbol{C}_2$ destroys all entanglement when it operates on a qubit, and p is proportional to the probability that $\boldsymbol{C}_2$ operates. The larger the p, the less entanglement there

is after qubits pass through the channel, thereby resulting in a lower F^*. The angle at which $\boldsymbol{C}_2$ rotates a qubit is given by $\arcsin(|\eta|)$. The larger $|\eta|$, the easier it is to detect which qubits are operated by $\boldsymbol{C}_2$, thereby resulting in a higher F^*. In particular, when $|\eta| = 1$, $F^* = 1$ provided that $p < 1$. Therefore, for amplitude-damping channels, it is possible to design recurrence QED algorithms that generate maximally entangled qubit pairs as long as the channel does not completely destroy entanglement. □

C. Achieving the optimal fidelity

The following algorithm first adapts to the channel so that the prepared qubit pairs have density matrices with a structure invariant to the channel. Then the algorithm employs recurrent operations to progressively improve the fidelity of the kept qubit pairs. These operations are specially designed to match the prepared density matrix structure, so that the proposed algorithm achieves the optimal fidelity in every round of distillation.

Algorithm 1. Adaptive recurrence QED algorithm.

(1) RSSP. For each qubit pair, the agents transform the density matrix into $\check{\boldsymbol{\rho}}$ using predistillation unitary operators $\boldsymbol{U}_{\mathrm{A}}$ and $\boldsymbol{U}_{\mathrm{B}}$.[4] Then Bob applies predistillation measurement operators

$$\boldsymbol{M}_{\mathrm{B}} = \begin{bmatrix} [1]\kappa & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{M}_{\bar{\mathrm{B}}} = \begin{bmatrix} \sqrt{1-\kappa^2} & 0 \\ 0 & 0 \end{bmatrix} \tag{20}$$

on his qubit, where $\kappa = \frac{\beta}{\alpha}$. If the measurement result corresponds to $\boldsymbol{M}_{\mathrm{B}}$, Bob performs no further action; otherwise, he notifies Alice via classical communication and the agents discard the qubit pair.

(2) First round distillation. The agents take two of the kept qubit pairs, perform the following operations, and repeat these operations on all kept qubit pairs.

(i) Each agent locally performs controlled-NOT (CNOT) operation, i.e., $\boldsymbol{U} = |00\rangle\langle 00| + |01\rangle\langle 01| + |10\rangle\langle 11| + |11\rangle\langle 10|$ on the two qubits at hand.

(ii) Each agent measures the target bit (i.e., the qubit in the second pair) using operators $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$, and transmits the measurement result to the other agent via classical communication.

(iii) If their measurement results do not agree, the agents discard the source qubit pair (i.e., the first pair). If the measurement results agree and correspond to $|1\rangle\langle 1|$, the agents keep the source qubit pair. If the measurement results agree and correspond to $|0\rangle\langle 0|$, the agents may choose to discard or keep the source qubit pair; the approach that discards or keeps the qubit pair in this case is referred to as the fidelity-prioritized (FP) or probability-prioritized (PP) approach, respectively.

(3) Following rounds. Agents perform the same operations as in the first round, except that they always adopt the PP approach, i.e., keep the source qubit pair as long as the measurement results agree. Repeat this step until the fidelity of the kept qubit pairs exceeds the required threshold. □

For notational convenience, denote the fidelity of the kept qubit pairs after the nth round of iteration as F_n, where $F_0 = F$. The following theorem characterizes the performance of the proposed algorithm.

Theorem 3. Performance of the proposed algorithm. After the RSSP and first round of distillation, a qubit pair is kept with probability

$$P_1 = \begin{cases} \frac{F_0^2\alpha^2\beta^4 + (1-F_0)^2\beta^2\gamma^2\delta^2}{2F_0\alpha^2\beta^2 + (1-F_0)(\alpha^2\gamma^2+\beta^2\delta^2)} & \text{for the FP approach} \\ \frac{4F_0^2\alpha^4\beta^4 + (1-F_0)^2(\alpha^2\gamma^2+\beta^2\delta^2)^2}{4F_0\alpha^4\beta^2 + 2(1-F_0)\alpha^2(\alpha^2\gamma^2+\beta^2\delta^2)} & \text{for the PP approach} \end{cases} \tag{21}$$

and fidelity

$$F_1 = \begin{cases} \frac{F_0^2}{F_0^2 + (1-F_0)^2(\frac{\gamma\delta}{\alpha\beta})^2} & \text{for the FP approach} \\ \frac{F_0^2}{F_0^2 + \frac{1}{4}(1-F_0)^2(\frac{\gamma^2}{\beta^2}+\frac{\delta^2}{\alpha^2})^2} & \text{for the PP approach} \end{cases}. \tag{22}$$

In the kth round ($k = 2,3,4,\ldots$) of distillation, a qubit pair is kept with probability

$$P_k = \tfrac{1}{2}\left[F_{k-1}^2 + (1-F_{k-1})^2\right] \tag{23}$$

and fidelity

$$F_k = \frac{F_{k-1}^2}{F_{k-1}^2 + (1-F_{k-1})^2}. \tag{24}$$

Proof. The proof is given in Appendix F. □

In the following, algorithms that adopt the FP and PP approaches in the first round of distillation are referred to as FP and PP algorithms, respectively.

Remark 4. Convergence speed of fidelity. For the FP algorithm, the density matrix of the kept qubit pair after the kth round of distillation is

$$\boldsymbol{\rho}^{(k)} = F_{k-1}|\Phi^+\rangle\langle\Phi^+| + (1-F_{k-1})|\Psi^+\rangle\langle\Psi^+|.$$

In this case, by comparing (18) with (22) or (24), it can be observed that the FP algorithm achieves the optimal fidelity in every round of distillation. This implies that the FP algorithm attains the fastest convergence speed with respect to the rounds of distillation.

The PP algorithm achieves a lower fidelity in the first round compared to the FP algorithm. On the other hand, (21) shows that the probability of keeping a qubit pair in the first round is higher with the PP algorithm compared to the FP algorithm by a factor more than 2. In particular, when the channel is phase damping, i.e., $\alpha = \beta = \gamma = \delta = \frac{1}{\sqrt{2}}$, the PP algorithm doubles the probability of keeping a qubit pair without lowering the fidelity achieved in the first round.

For the first recurrence QED algorithm (which will be referred to as the BBPSSW algorithm in this paper) proposed in [19], the fidelity of the kept qubit pairs after the kth round of distillation is given by

$$F_k = \frac{F_{k-1}^2 + \frac{1}{9}(1-F_{k-1})^2}{F_{k-1}^2 + \frac{2}{3}F_{k-1}(1-F_{k-1}) + \frac{5}{9}(1-F_{k-1})^2}. \tag{25}$$

[4]Given a TKO channel, if $\eta = 0$, $\boldsymbol{U}_{\mathrm{A}}$ and $\boldsymbol{U}_{\mathrm{B}}$ are determined by (B3). Otherwise, one can obtain $\boldsymbol{\rho}$ via (3), then perform singular value decomposition and Schmidt decomposition sequentially to get (B7) and (B8), and finally determine $\boldsymbol{U}_{\mathrm{A}}$ and $\boldsymbol{U}_{\mathrm{B}}$ via (B16).

Therefore, when $F_0 > \frac{1}{2}$, it can be shown using (25) that

$$\lim_{k\to\infty} \frac{1-F_k}{1-F_{k-1}} = \frac{2}{3}. \tag{26}$$

For the proposed algorithms, when $F_0 > \frac{1}{2}$, it can be shown using (24) that

$$\lim_{k\to\infty} \frac{1-F_k}{1-F_{k-1}} = 0, \quad \lim_{k\to\infty} \frac{1-F_k}{(1-F_{k-1})^2} = 1. \tag{27}$$

Equation (26) shows that with the BBPSSW algorithm the fidelity of the qubit pairs converges to 1 linearly at rate $\frac{2}{3}$, whereas (27) shows that with the proposed algorithms the fidelity converges to 1 quadratically. Hence, the convergence speed of the proposed algorithms is significantly improved, i.e., from linear to quadratic, compared to the BBPSSW algorithm. □

Remark 5. Connection to the QPA algorithm. When the channel is phase damping, i.e., $\eta = 0$, (i) the predistillation unitary operators $\boldsymbol{U}_{\mathrm{A}} = \boldsymbol{U}_{\mathrm{B}} = \boldsymbol{H}$ according to (B3) and (ii) the predistillation measurement operator $\boldsymbol{M}_{\mathrm{B}} = \mathbb{I}_2$ since $\alpha = \beta$ according to (10). In this case, both local operators employed by Alice and Bob in the RSSP are equal to the Hadamard transform $\boldsymbol{H}$, and hence the PP algorithm becomes the QPA algorithm in [20]. Therefore, the QPA algorithm is a special case of the PP algorithm, which employs fixed predistillation operators for all channels. With such nonadaptive predistillation operators, the convergence of the fidelity achieved by the QPA algorithm is not guaranteed [20]. With the proposed adaptive predistillation operators, the fidelity achieved by both FP and PP algorithms converges quadratically for TKO channels. The proposed algorithms may be applied to more general channels, yet their convergence property for such channels remains to be characterized. □

Remark 6. Benefit of channel adaptation. In the proposed algorithms, the channel adaptation takes place in the RSSP. As shown in Theorem 3 and Remark 4, despite its simplicity of involving single-qubit operations only in the initial step, RSSP is the keystone to improve the effectiveness of distillation for TKO channels. With the BBPSSW algorithm [19], in addition to the distillation operations, random bilateral rotations are required to restore the desired density matrix structure in every round of distillation. With the QPA algorithm, no random rotations are required, yet the density matrix structure may not be preserved for different rounds of distillation. In the proposed algorithms, the RSSP adapts to the channel so that the prepared qubit pairs have density matrices with a structure invariant to the channel. As a result, the distillation operation itself, which involves only the CNOT operation and single-qubit measurements, is sufficient to maintain the density matrix structure in every round of distillation. This feature enables a simple QED algorithm with guaranteed convergence. Hence, channel adaptation also improves the implementability of QED algorithms. □

IV. NUMERICAL RESULTS

This section provides numerical results to demonstrate the performance of the proposed algorithms. In particular, the proposed FP and PP algorithms are compared with the BBPSSW algorithm in [19] and the QPA in [20] for a required fidelity $F_{\mathrm{th}} = 0.99$.

Figure 3 shows the fidelity of kept qubit pairs as a function of the rounds of distillation for three types of channels, i.e., a phase-damping channel, a "midpoint" channel,[5] and an amplitude-damping channel. When the channel is phase damping, the fidelities achieved by FP, PP, and QPA algorithms are the same, which is consistent with the observations made in Remarks 4 and 5. When the channel is midpoint or amplitude damping, the QPA algorithm does not achieve the required fidelity, illustrating its converge issue. The FP, PP, and the BBPSSW algorithm achieve the required fidelity on all channels, with the proposed algorithms using much less rounds of distillation. For instance, when the channel is amplitude damping, the BBPSSW algorithm requires 24 rounds of distillation, whereas the FP and PP algorithms only require one and three rounds, respectively. Since the yield is reduced by at least half after each round of distillation, the yield of the proposed algorithms is significantly higher than the classical one for all the considered channels.

Figures 4 and 5 show the yield of the distillation algorithms as a function of the noise severity p and the channel type parameter $|\eta|$, respectively. As a benchmark, the bound on distillable entanglement [49,50] is also plotted in these figures. While the achievability of this bound remains unknown, it is arguably the best known upper bound on the yield of any QED algorithms [51].

Figure 4 shows that while the yield of all algorithms decreases with noise severity p, the proposed algorithms are much more resilient to noise compared to the BBPSSW algorithm. For instance, when $p = 0.4$, the yield of the proposed algorithms is about 360 times higher than that of the BBPSSW algorithm and is only 1.6 times away from the best known upper bound. Comparing the two proposed algorithms, the FP algorithm performs better for large p, whereas the PP algorithm performs better for small p. This shows that when the noise is severe it is beneficial to increase the achieved fidelity at a cost of reducing the probability of keeping qubit pairs.

Figure 5 shows that the proposed algorithms are resilient to the variance of the channel-type parameter η. The yield of the better-performed proposed algorithm is at least 110 times higher than that of the BBPSSW algorithm and is only 0.6 times (when $\eta = 0$) to 1.6 times (when $|\eta| = 1$) away from the best known upper bound. Given that the proposed algorithms only require quantum operations on one or two qubits, the gap between the yield of the proposed algorithms and the upper bound indicates that the proposed algorithms achieve a desirable balance between efficiency and implementability. For the QPA algorithm, it can be seen that this algorithm has the same efficiency as the PP algorithm when the channel is phase damping, which is consistent with Remark 5. Yet the QPA algorithm does not achieve the required fidelity when $\arcsin|\eta| \geqslant 0.04\pi$. This illustrates the importance of channel adaptation. Comparing the two proposed algorithms, the FP algorithm is more efficient when the channel tends towards an

[5]This channel [$\arcsin(|\eta|) = \frac{\pi}{4}$] can be thought of as the midpoint of phase-damping channels [$\arcsin(|\eta|) = 0$] and amplitude-damping channels [$\arcsin(|\eta|) = \frac{\pi}{2}$].

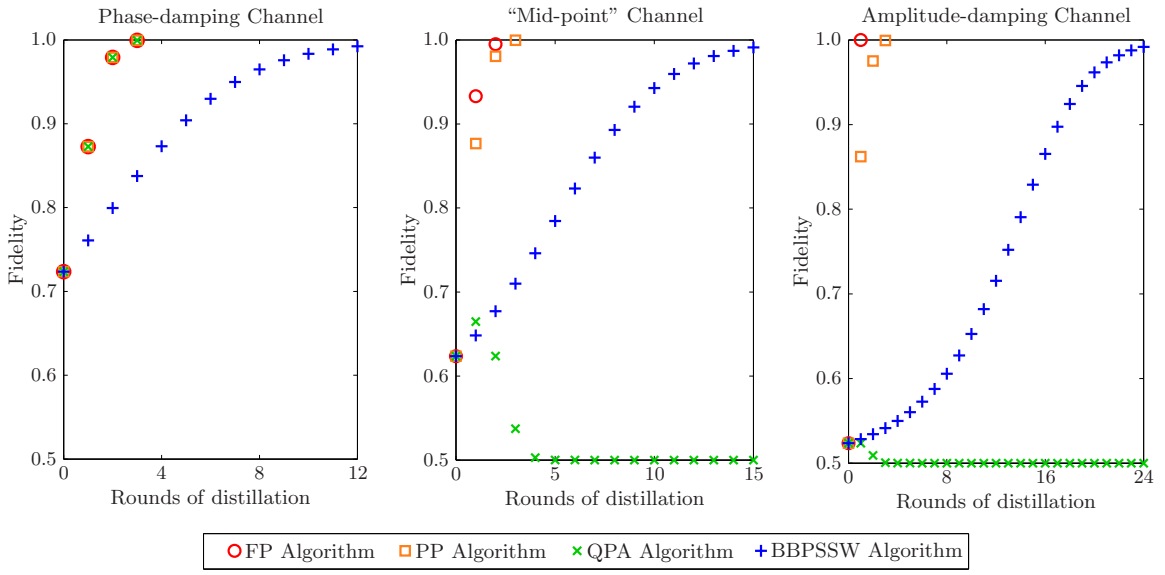

FIG. 3. The achieved fidelity as a function of the rounds of distillation for a phase-damping channel, a midpoint channel, and an amplitude-damping channel. For all channels, the noise severity parameter $p = 0.8$.

amplitude-damping channel (i.e., $|\eta|$ approaches 1), and the PP algorithm is more efficient when the channel tends towards a phase-damping channel (i.e., $|\eta|$ approaches zero). This is consistent with Theorems 1 and 3, which show that the benefit of increasing the fidelity by adopting the FP algorithm is greater when η is close to 1 and vice versa.

Finally, the performance of recurrence QED algorithms with imperfect operations is evaluated. Noticing that in practice, the error rates of two-qubit operations are typically much higher than those of one-qubit operations [40], we focus on the effect of imperfect CNOT operations. Adopting the error model used in [52,53], Fig. 6 shows the maximum achievable fidelity of various algorithms as a function of the error rate of CNOT operations. It can be seen that while the maximum achievable fidelity of all algorithms drops when the error rate increases, the proposed algorithms, particularly the FP algorithm, are less sensitive to imperfect CNOT operations. This is because the proposed algorithms require fewer rounds of distillation, and consequently fewer CNOT operations, to achieve a certain fidelity; this reduces the overall effect of CNOT imperfection on the maximum achievable fidelity.

V. CONCLUSION

Among various types of QED algorithms, the recurrence ones require quantum operations on the minimum number of qubits and can generate maximally entangled qubit pairs even when the noise in the channel is severe. Despite their advantages, the efficiency issue of recurrence QED algorithms has

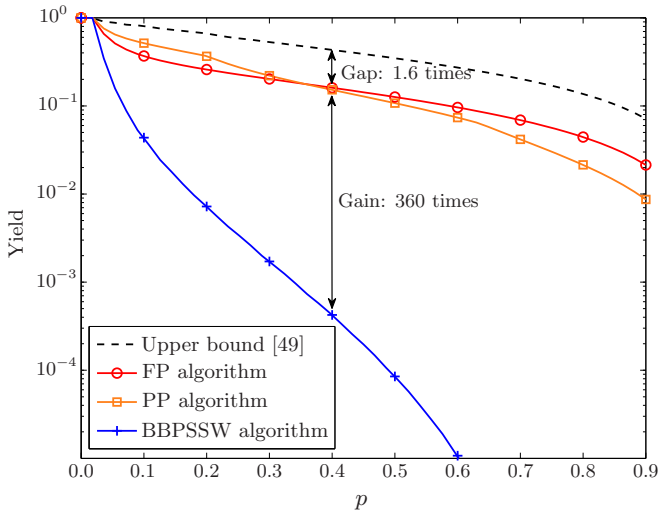

FIG. 4. The efficiency of different recurrence QED algorithms as a function of noise severity p for amplitude-damping channels ($p \in [0,0.9]$, $\eta = 1$). The QPA algorithm is not plotted as it does not achieve the required fidelity on amplitude-damping channels.

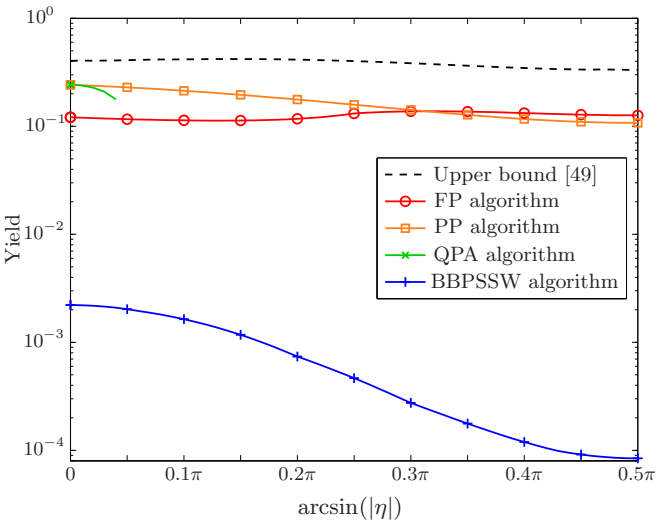

FIG. 5. The efficiency of the two proposed QED algorithms as a function of channel type parameter $|\eta|$ ($p = 0.5$, $|\eta| \in [0,1]$).

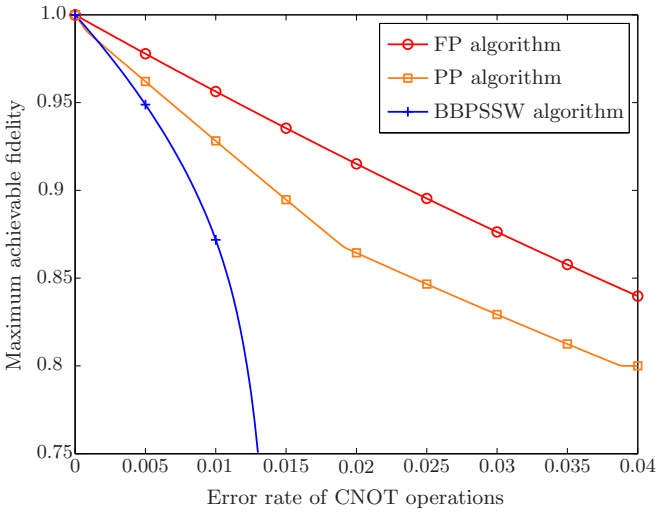

FIG. 6. The maximum achievable fidelity of different algorithms as a function of the error rate of CNOT operations. In this figure, $p = 0.5$, $\eta = 1$. The QPA algorithm is not plotted as it does not achieve the required fidelity on amplitude-damping channels.

not been thoroughly investigated in the literature. In this paper, we first characterize the effect of a single-qubit TKO channel on the entanglement of a qubit pair shared by the agents via this channel. We then determine the optimal fidelity that can be achieved by performing LOCC on two such qubit pairs. Finally we propose two adaptive recurrence QED algorithms, one of which achieves the optimal fidelity. The proposed algorithms preserve the density matrix structure in every round of distillation, avoiding the need of additional random rotations. This enables simple QED algorithms with guaranteed convergence for TKO channels. In fact, the convergence speed of both algorithms is improved from linear to quadratic compared to the BBPSSW algorithm. Numerical results confirm that the proposed algorithms significantly improve the efficiency of recurrence QED algorithms. These results also indicate that the benefit of achieving optimal fidelity is greater when the noise is severe, or the channel tends towards an amplitude-damping channel.

ACKNOWLEDGMENTS

The authors would like to thank Aram Harrow and Peter W. Shor for the valuable discussions and suggestions. This research was supported in part by the MIT Institute for Soldier Nanotechnologies.

APPENDIX A: PROOF OF LEMMA 1

Consider a single-qubit TKO channel represented by $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$. We will first prove the theorem for the case in which $\text{rank}\{\boldsymbol{C}_2\} = 1$, then show that the case in which $\text{rank}\{\boldsymbol{C}_2\} = 2$ can be transformed into the prior case.

When $\text{rank}\{\boldsymbol{C}_2\} = 1$, singular value decomposition (SVD) of $\boldsymbol{C}_2$ shows that there exist $|i\rangle, |j\rangle \in \mathbb{C}^2$, $p \in (0,1]$, and $\epsilon \in \mathbb{R}$ such that

$$\boldsymbol{C}_2 = \sqrt{p}e^{\imath\epsilon}|i\rangle\langle j|. \tag{A1}$$

Noting that a quantum operator is invariant up to an overall phase change, ϵ can be any real number. Recall from (2) that

$$\boldsymbol{C}_1^\dagger \boldsymbol{C}_1 = \mathbb{I}_2 - \boldsymbol{C}_2^\dagger \boldsymbol{C}_2. \tag{A2}$$

Substituting the SVD of $\boldsymbol{C}_1 = \boldsymbol{U}_\text{c}\boldsymbol{D}_\text{c}\boldsymbol{V}_\text{c}^\dagger$ and (A1) into (A2), one can get

$$\boldsymbol{V}_\text{c}\boldsymbol{D}_\text{c}^2\boldsymbol{V}_\text{c}^\dagger = |\tilde{j}\rangle\langle\tilde{j}| + (1-p)|j\rangle\langle j|, \tag{A3}$$

where $\langle j|\tilde{j}\rangle = 0$. Since $\boldsymbol{D}_\text{c}^2$ is diagonal and $\boldsymbol{V}_\text{c}$ is unitary,

$$\boldsymbol{D}_\text{c} = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{bmatrix}, \quad \boldsymbol{V}_\text{c} = [|\tilde{j}\rangle\ |j\rangle].$$

Hence, there exists $|\tilde{k}\rangle$ and $|k\rangle$ with $\langle\tilde{k}|k\rangle = 0$ such that

$$\boldsymbol{C}_1 = |\tilde{k}\rangle\langle\tilde{j}| + \sqrt{1-p}|k\rangle\langle j|. \tag{A4}$$

It can be verified that $\boldsymbol{C}_1$ in (A4) and $\boldsymbol{C}_2$ in (A1) can be expressed in the form given in (4), with

$$\begin{aligned} &\boldsymbol{U} = |\tilde{k}\rangle\langle 0| + |k\rangle\langle 1|, \quad \boldsymbol{V} = |\tilde{j}\rangle\langle 0| + |j\rangle\langle 1|, \\ &\eta = e^{\imath\epsilon}\langle\tilde{k}|i\rangle, \quad \zeta = e^{\imath\epsilon}\langle k|i\rangle, \\ &\epsilon = -\text{pha}\{\langle k|i\rangle\}. \end{aligned} \tag{A5}$$

This completes the proof for the case with $\text{rank}\{\boldsymbol{C}_2\} = 1$.

Now consider the case in which $\text{rank}\{\boldsymbol{C}_2\} = 2$. Since $\boldsymbol{C}_2$ is full rank, $\det\{\boldsymbol{C}_2\} \neq 0$. Consider the equation

$$\det\{-\boldsymbol{C}_1 + x\boldsymbol{C}_2\} = 0. \tag{A6}$$

This is a second-order polynomial equation of x, for which the coefficient of the second-order term is $\det\{\boldsymbol{C}_2\} \neq 0$. Therefore, the fundamental theorem of algebra implies that (A6) must have at least one solution. Denote x_0 as one of the solutions of (A6). Recall from [44], Sec 3.3] that any single-qubit TKO channel with operators $\{\boldsymbol{C}_k\}$ can be equivalently represented by operators $\{\tilde{\boldsymbol{C}}_k\}$ satisfying

$$[\tilde{\boldsymbol{C}}_1\ \tilde{\boldsymbol{C}}_2] = [\boldsymbol{C}_1\ \boldsymbol{C}_2](\boldsymbol{A} \otimes \mathbb{I}_2), \tag{A7}$$

where $\boldsymbol{A}$ is an arbitrary unitary matrix. In particular, let

$$\boldsymbol{A} = \frac{1}{\sqrt{1+|x_0|^2}}\begin{bmatrix} x_0^\dagger & -1 \\ 1 & x_0 \end{bmatrix},$$

then $\det\{\tilde{\boldsymbol{C}}_2\} = \frac{\det\{-\boldsymbol{C}_1 + x_0\boldsymbol{C}_2\}}{1+|x_0|^2} = 0$. Thus $\text{rank}\{\tilde{\boldsymbol{C}}_2\} \leqslant 1$. If it were the case that $\text{rank}\{\tilde{\boldsymbol{C}}_2\} = 0$, then $\boldsymbol{C}_2 = \boldsymbol{0}$ implying that the channel has only one operator. This contradicts that the channel has two operators. Hence, $\text{rank}\{\tilde{\boldsymbol{C}}_2\} = 1$, which is the case that has been proven above.

APPENDIX B: PROOF OF THEOREM 1

Since $\text{rank}\{\boldsymbol{\rho}_0\} = 1$, and the channel has only two operators, from (3), $\text{rank}\{\boldsymbol{\rho}\} \leqslant 2$. Since density matrices are Hermitian, the spectral decomposition gives

$$\boldsymbol{\rho} = F|\psi\rangle\langle\psi| + (1-F)|\phi\rangle\langle\phi|, \tag{B1}$$

where

$$\langle\psi|\phi\rangle = 0. \tag{B2}$$

In (B1), we have used the fact that density matrices have trace 1. Without loss of generality, assume $F \in [\frac{1}{2}, 1]$.

If $p = 0$, (3) and (5) imply that $\boldsymbol{\rho} = \boldsymbol{\rho}_0$. Setting $\boldsymbol{U}_{\mathrm{A}} = \boldsymbol{U}_{\mathrm{B}} = \mathbb{I}_2$ in (6), it is straightforward that the theorem holds. Also, if $\eta = 0$, the channel is phase damping. Then

$$\begin{aligned}\boldsymbol{\rho} &= \frac{1}{2}[(|00\rangle + \sqrt{1-p}|11\rangle)(\langle 00| \\ &\quad + \sqrt{1-p}\langle 11| + p|11\rangle\langle 11|] \\ &= \frac{1}{2}(|00\rangle\langle 00| + \sqrt{1-p}|00\rangle\langle 11| \\ &\quad + \sqrt{1-p}|11\rangle\langle 00| + |11\rangle\langle 11|) \\ &= F|\psi\rangle\langle\psi| + (1-F)|\phi\rangle\langle\phi|,\end{aligned}$$

where

$$\begin{aligned}F &= \frac{1+\sqrt{1-p}}{2}, \\ |\psi\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \\ |\phi\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle).\end{aligned}$$

Setting both local unitary operators in (6) to be Hadamard transform, i.e.,

$$\boldsymbol{U}_{\mathrm{A}} = \boldsymbol{U}_{\mathrm{B}} = \boldsymbol{H} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}, \tag{B3}$$

it is easy to see that the theorem also holds for the case of $\eta = 0$. Therefore, the following analysis considers the case for which $p \in (0,1)$ and $|\eta| > 0$.

We first determine the value of F. Set $\boldsymbol{A} = \left[\begin{smallmatrix}\kappa^\dagger & -\lambda^\dagger\\ \lambda & \kappa\end{smallmatrix}\right]$ in (A7) with $\kappa,\lambda \in \mathbb{C}$ such that $|\kappa|^2 + |\lambda|^2 = 1$. Then, it can be shown that

$$\begin{aligned}\boldsymbol{\rho} &\overset{(a)}{=} \sum_{k=1}^{2}(\mathbb{I}_2 \otimes \tilde{\boldsymbol{C}}_k)|\Phi^+\rangle\langle\Phi^+|(\mathbb{I}_2 \otimes \tilde{\boldsymbol{C}}_k)^\dagger \\ &\overset{(b)}{=} [\mathbb{I}_2 \otimes (\kappa^\dagger \boldsymbol{C}_1 + \lambda \boldsymbol{C}_2)]|\Phi^+\rangle\langle\Phi^+| \\ &\quad \times [\mathbb{I}_2 \otimes (\kappa^\dagger \boldsymbol{C}_1 + \lambda \boldsymbol{C}_2)]^\dagger \\ &\quad + [\mathbb{I}_2 \otimes (-\lambda^\dagger \boldsymbol{C}_1 + \kappa \boldsymbol{C}_2)]|\Phi^+\rangle\langle\Phi^+| \\ &\quad \times [\mathbb{I}_2 \otimes (-\lambda^\dagger \boldsymbol{C}_1 + \kappa \boldsymbol{C}_2)]^\dagger \\ &\overset{(c)}{=} \frac{1}{2}(\boldsymbol{v}_1\boldsymbol{v}_1^\dagger + \boldsymbol{v}_2\boldsymbol{v}_2^\dagger),\end{aligned} \tag{B4}$$

where

$$\boldsymbol{v}_1 = \begin{bmatrix}\kappa^\dagger \\ 0 \\ \lambda\eta\sqrt{p} \\ \kappa^\dagger\sqrt{1-p} + \lambda\zeta\sqrt{p}\end{bmatrix},$$

$$\boldsymbol{v}_2 = \begin{bmatrix}-\lambda^\dagger \\ 0 \\ \kappa\eta\sqrt{p} \\ -\lambda^\dagger\sqrt{1-p} + \kappa\zeta\sqrt{p}\end{bmatrix}.$$

In (B4), (a) is due to (3) together with the equivalence between $\{\boldsymbol{C}_k\}$ and $\{\tilde{\boldsymbol{C}}_k\}$, (b) is due to (A7), and (c) is due to (5).

In order to make the last line of (B4) a spectral decomposition of $\boldsymbol{\rho}$, it is necessary to make $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ orthogonal. A sufficient condition for $\boldsymbol{v}_1^\dagger\boldsymbol{v}_2 = 0$ is given by

$$\kappa = \left(\frac{\sqrt{\frac{1-p}{1-|\eta|^2 p}}+1}{2}\right)^{\frac{1}{2}}, \quad \lambda = \sqrt{1-\kappa^2}.$$

Setting $|\psi\rangle = \frac{\boldsymbol{v}_1}{||\boldsymbol{v}_1||}$ and $|\phi\rangle = \frac{\boldsymbol{v}_2}{||\boldsymbol{v}_2||}$ in the last line of (B4), one can get

$$\boldsymbol{\rho} = \tfrac{1}{2}||\boldsymbol{v}_1||^2|\psi\rangle\langle\psi| + \tfrac{1}{2}||\boldsymbol{v}_2||^2|\phi\rangle\langle\phi|, \tag{B5}$$

which is a spectral decomposition of $\boldsymbol{\rho}$. Since the spectrum of a matrix is unique, by comparing (B5) with (B1), one gets

$$F = \tfrac{1}{2}||\boldsymbol{v}_1||^2 = \tfrac{1}{2} + \tfrac{1}{2}\sqrt{(1-p)(1-|\eta|^2 p)}, \tag{B6}$$

which proves (9).

We next show (7) and (8). Perform Schmidt decomposition on the eigenvectors of $\boldsymbol{\rho}$ as

$$|\psi\rangle = \alpha|wx\rangle + \beta|\tilde{w}\tilde{x}\rangle, \tag{B7}$$

$$|\phi\rangle = \gamma|yz\rangle + \delta|\tilde{y}\tilde{z}\rangle, \tag{B8}$$

where $\langle s|\tilde{s}\rangle = 0$ for $s \in \{w,x,y,z\}$, and $\alpha,\beta,\gamma,\delta \in [0,1]$ satisfying

$$\alpha^2 + \beta^2 = \gamma^2 + \delta^2 = 1. \tag{B9}$$

Without loss of generality, assume $\beta \leqslant \alpha$, $\gamma \leqslant \delta$.

From (B6), when $p \in (0,1)$, $F \in (\frac{1}{2},1)$. Substituting (5) into (3) and taking the partial trace over different qubits, one can obtain the density matrices of the first and second qubits, i.e.,

$$\begin{aligned}\boldsymbol{\rho}_1 &= \mathrm{tr}_2\{\boldsymbol{\rho}\} = \frac{1}{2}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \\ \boldsymbol{\rho}_2 &= \mathrm{tr}_1\{\boldsymbol{\rho}\} = \frac{1}{2}\begin{bmatrix}1+|\eta|^2 p & \eta\zeta p\\ \eta^\dagger\zeta p & 1-p+\zeta^2 p\end{bmatrix}.\end{aligned} \tag{B10}$$

On the other hand, substituting (B7) and (B8) into (B1) and taking the partial trace, one can obtain an alternative expression of $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ in terms of $|x\rangle$, $x \in \{a,b,c,d\}$. This together with (B10) gives

$$\begin{aligned}&F(\alpha^2|w\rangle\langle w| + \beta^2|\tilde{w}\rangle\langle\tilde{w}|) + (1-F)(\gamma^2|y\rangle\langle y| + \delta^2|\tilde{y}\rangle\langle\tilde{y}|) \\ &\quad = \frac{1}{2}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},\end{aligned} \tag{B11}$$

$$\begin{aligned}&F(\alpha^2|x\rangle\langle x| + \beta^2|\tilde{x}\rangle\langle\tilde{x}|) + (1-F)(\gamma^2|z\rangle\langle z| + \delta^2|\tilde{z}\rangle\langle\tilde{z}|) \\ &\quad = \frac{1}{2}\begin{bmatrix}1+|\eta|^2 p & \eta\zeta p\\ \eta^\dagger\zeta p & 1-p+\zeta^2 p\end{bmatrix}.\end{aligned} \tag{B12}$$

We claim that $\gamma < \delta$ when $|\eta| > 0$. If it were not the case, then $\gamma = \delta = \frac{1}{\sqrt{2}}$. Thus

$$\gamma^2|y\rangle\langle y| + \delta^2|\tilde{y}\rangle\langle\tilde{y}| = \frac{1}{2}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} \tag{B13}$$

because $\langle c|\tilde{c}\rangle = 0$. Substituting (B13) into the left side of (B11), one can get

$$\alpha^2|w\rangle\langle w| + \beta^2|\tilde{w}\rangle\langle\tilde{w}| = \frac{1}{2}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}. \tag{B14}$$

Since $\langle w|\tilde{w}\rangle = 0$, the left side of (B14) is a spectral decomposition of the right side, implying $\alpha = \beta = \frac{1}{\sqrt{2}}$. Substituting

$\alpha=\beta=\gamma=\delta=\frac{1}{\sqrt{2}}$ into the left side of (B12), and since $\langle x|\tilde{x}\rangle=\langle z|\tilde{z}\rangle=0$, one can get

$$\begin{bmatrix}1&0\\0&1\end{bmatrix}=\begin{bmatrix}1+|\eta|^2p&\eta\zeta p\\ \eta^\dagger\zeta p&1-p+\zeta^2p\end{bmatrix}, \tag{B15}$$

which holds only if $|\eta|=0$. This contradicts the fact that $|\eta|>0$ and thus proves the claim.

Next construct two unitary operators as

$$\boldsymbol{U}_{\mathrm{A}}=|0\rangle\langle w|+|1\rangle\langle\tilde{w}|,\quad \boldsymbol{U}_{\mathrm{B}}=|0\rangle\langle x|+|1\rangle\langle\tilde{x}|. \tag{B16}$$

Substituting this into (B1), it can be obtained that

$$\breve{\boldsymbol{\rho}}=F|\mu\rangle\langle\mu|+(1-F)|\nu\rangle\langle\nu|, \tag{B17}$$

in which

$$|\mu\rangle=\alpha|00\rangle+\beta|11\rangle, \tag{B18}$$

$$|\nu\rangle=\gamma|y_{\mathrm{r}}z_{\mathrm{r}}\rangle+\delta|\tilde{y}_{\mathrm{r}}\tilde{z}_{\mathrm{r}}\rangle, \tag{B19}$$

where the ket notations with subscript "r" denote the rotated version of the original ones, e.g., $|y_{\mathrm{r}}\rangle=\boldsymbol{U}_{\mathrm{A}}|y\rangle$ and $|\tilde{z}_{\mathrm{r}}\rangle=\boldsymbol{U}_{\mathrm{B}}|\tilde{z}\rangle$. Equation (B17) gives the structure of (6), and (B18) proves (7).

The following analysis focuses on proving (8). Since $\boldsymbol{U}_{\mathrm{A}}$ and $\boldsymbol{U}_{\mathrm{B}}$ are unitary, (B2) implies $\langle\mu|\nu\rangle=0$, which gives

$$\alpha\gamma\langle00|y_{\mathrm{r}}z_{\mathrm{r}}\rangle+\alpha\delta\langle00|\tilde{y}_{\mathrm{r}}\tilde{z}_{\mathrm{r}}\rangle+\beta\gamma\langle11|y_{\mathrm{r}}z_{\mathrm{r}}\rangle+\beta\delta\langle11|\tilde{y}_{\mathrm{r}}\tilde{z}_{\mathrm{r}}\rangle=0. \tag{B20}$$

Substituting $|w\rangle=\boldsymbol{U}_{\mathrm{A}}^\dagger|0\rangle$, $|\tilde{w}\rangle=\boldsymbol{U}_{\mathrm{A}}^\dagger|1\rangle$, $|y\rangle=\boldsymbol{U}_{\mathrm{A}}^\dagger|y_{\mathrm{r}}\rangle$, and $|\tilde{y}\rangle=\boldsymbol{U}_{\mathrm{A}}^\dagger|\tilde{y}_{\mathrm{r}}\rangle$ into (B11), one can get

$$\begin{aligned}&F(\alpha^2|0\rangle\langle0|+\beta^2|1\rangle\langle1|)+(1-F)(\gamma^2|y_{\mathrm{r}}\rangle\langle y_{\mathrm{r}}|+\delta^2|\tilde{y}_{\mathrm{r}}\rangle\langle\tilde{y}_{\mathrm{r}}|)\\&\quad=\tfrac{1}{2}(|0\rangle\langle0|+|1\rangle\langle1|).\end{aligned} \tag{B21}$$

Since $\boldsymbol{U}_{\mathrm{A}}$ and $\boldsymbol{U}_{\mathrm{B}}$ are unitary and $\langle s|\tilde{s}\rangle=0$, $\langle s_{\mathrm{r}}|\tilde{s}_{\mathrm{r}}\rangle=0$, where $s\in\{y,z\}$. Since $|0\rangle$, $|1\rangle$ and $|y_{\mathrm{r}}\rangle$, $|\tilde{y}_{\mathrm{r}}\rangle$ are two sets of orthonormal bases for two-dimensional Hilbert space, there exist $a,b\in\mathbb{C}$, $|a|^2+|b|^2=1$, such that

$$|y_{\mathrm{r}}\rangle=a|0\rangle+b|1\rangle,\quad |\tilde{y}_{\mathrm{r}}\rangle=-b^\dagger|0\rangle+a^\dagger|1\rangle. \tag{B22}$$

Substitute (B22) into (B21), then

$$\begin{aligned}&\big(F\alpha^2+(1-F)(\gamma^2|a|^2+\delta^2|b|^2)-\tfrac{1}{2}\big)|0\rangle\langle0|\\&+\big(F\beta^2+(1-F)(\gamma^2|b|^2+\delta^2|a|^2)-\tfrac{1}{2}\big)|1\rangle\langle1|\\&+(1-F)(\gamma^2-\delta^2)ab^\dagger|0\rangle\langle1|\\&+(1-F)(\gamma^2-\delta^2)a^\dagger b|1\rangle\langle0|=\boldsymbol{0}.\end{aligned} \tag{B23}$$

Therefore

$$F\alpha^2+(1-F)(\gamma^2|a|^2+\delta^2|b|^2)-\tfrac{1}{2}=0, \tag{B24}$$

$$F\beta^2+(1-F)(\gamma^2|b|^2+\delta^2|a|^2)-\tfrac{1}{2}=0, \tag{B25}$$

$$(1-F)(\gamma^2-\delta^2)ab^\dagger=0, \tag{B26}$$

$$(1-F)(\gamma^2-\delta^2)a^\dagger b=0. \tag{B27}$$

Since $F<1$ and $\gamma<\delta$, from (B26) and (B27), one can get $a=0$ or $b=0$. Without loss of generality, let $b=0$, then $|a|=1$. Therefore (B22) becomes

$$|y_{\mathrm{r}}\rangle=e^{\imath\theta_{\mathrm{a}}}|0\rangle,\quad |\tilde{y}_{\mathrm{r}}\rangle=e^{-\imath\theta_{\mathrm{a}}}|1\rangle, \tag{B28}$$

where $\theta_{\mathrm{a}}=\mathrm{pha}\{a\}$. Substituting (B28) into (B24) and (B25) gives

$$F\alpha^2+(1-F)\gamma^2=F\beta^2+(1-F)\delta^2=\tfrac{1}{2}. \tag{B29}$$

Since $F>\frac{1}{2}$, substituting (B9) into (B29) shows that

$$\gamma<\beta<\frac{\sqrt{2}}{2}<\alpha<\delta. \tag{B30}$$

Moreover, substituting (B28) into (B20) gives

$$e^{\imath\theta_{\mathrm{a}}}\alpha\gamma\langle0|z_{\mathrm{r}}\rangle+e^{-\imath\theta_{\mathrm{a}}}\beta\delta\langle1|\tilde{z}_{\mathrm{r}}\rangle=0,$$

which implies

$$|\alpha\gamma\langle0|z_{\mathrm{r}}\rangle|=|\beta\delta\langle1|\tilde{z}_{\mathrm{r}}\rangle|. \tag{B31}$$

On the other hand, since $\langle z_{\mathrm{r}}|\tilde{z}_{\mathrm{r}}\rangle=0$, it can be verified that $|\langle0|z_{\mathrm{r}}\rangle|=|\langle1|\tilde{z}_{\mathrm{r}}\rangle|$. Therefore from (B30), $|\alpha\gamma\langle0|z_{\mathrm{r}}\rangle|\leqslant|\beta\delta\langle1|\tilde{z}_{\mathrm{r}}\rangle|$, where the equality holds only if $|\langle0|z_{\mathrm{r}}\rangle|=|\langle1|\tilde{z}_{\mathrm{r}}\rangle|=0$. This result together with (B31) implies $|z_{\mathrm{r}}\rangle=e^{\imath\theta_z}|1\rangle$, $|\tilde{z}_{\mathrm{r}}\rangle=e^{\imath\theta_{\tilde{z}}}|0\rangle$, for some $\theta_z,\theta_{\tilde{z}}\in[0,2\pi)$. Further noting that a quantum state is invariant up to an overall phase change, one can get

$$|\nu\rangle=\gamma|01\rangle+\delta e^{\imath\theta}|10\rangle, \tag{B32}$$

where $\theta=\theta_{\tilde{z}}-\theta_z-2\theta_{\mathrm{a}}$. Equation (B32) proves (8). Therefore the local unitary operators $\boldsymbol{U}_{\mathrm{A}}$ and $\boldsymbol{U}_{\mathrm{B}}$ exhibited in (B16) give (6)–(8).

Finally, we show that α, β, γ, and δ satisfy (10) and (11). Substitute (B18) and (B32) into (B17), then the density matrix of the second qubit $\breve{\boldsymbol{\rho}}_2=\mathrm{tr}_1\{\breve{\boldsymbol{\rho}}\}$ becomes

$$\begin{aligned}\breve{\boldsymbol{\rho}}_2=&[F\alpha^2+(1-F)\delta^2]|0\rangle\langle0|\\&+[F\beta^2+(1-F)\gamma^2]|1\rangle\langle1|.\end{aligned} \tag{B33}$$

Noting that unitary operations do not change the determinant of a matrix, $\det\{\breve{\boldsymbol{\rho}}_2\}=\det\{\boldsymbol{\rho}_2\}$. Therefore, from (B10) and (B33) one can get

$$\begin{aligned}&[F\alpha^2+(1-F)\delta^2][F\beta^2+(1-F)\gamma^2]\\&=\tfrac{1}{4}[(1+|\eta|^2p)(1-p+\zeta^2p)-|\eta\zeta p|^2].\end{aligned} \tag{B34}$$

Substituting (B9) and (B29) into (B34), one can get

$$\alpha=\sqrt{\frac{1}{2}+\frac{|\eta|p}{4F}},\quad \delta=\sqrt{\frac{1}{2}+\frac{|\eta|p}{4(1-F)}}.$$

This completes the proof.

APPENDIX C: PROOF OF LEMMA 2

Equation (13) holds trivially when $\gamma=0$ since fidelity of any qubit pair cannot exceed 1, i.e., $f(F,\alpha,\beta,\gamma,\delta,\theta)\leqslant1$. Hence, it remains to consider the case for which $0<\gamma\leqslant\delta<1$, which will be proved by contradiction. Suppose Lemma 2

is false, then $\forall F \in (\frac{1}{2},1]$,

$$f\left(F,\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},0\right) \leqslant \frac{F^2}{F^2+(1-F)^2}, \tag{C1}$$

but there exists some F_0, α_0, β_0, γ_0, δ_0, and θ_0 such that

$$f(F_0,\alpha_0,\beta_0,\gamma_0,\delta_0,\theta_0) > \frac{F_0^2}{F_0^2+(1-F_0)^2\left(\frac{\gamma_0\delta_0}{\alpha_0\beta_0}\right)^2}. \tag{C2}$$

Then contradiction would arise if there exist some $\tilde{F} \in (\frac{1}{2},1]$ such that (C1) does not hold. To show the existence of such $\tilde{F}$, the RSSP method is employed to transform a given density matrix with parameters $F=\tilde{F}$, $\alpha=\beta=\gamma=\delta=\frac{1}{\sqrt{2}}$, and $\theta=0$ to another density matrix with parameters F_0, α_0, β_0, γ_0, δ_0, and θ_0 via LOCC. In particular, consider that Alice measures her qubit using local operators

$$\boldsymbol{M}_{\mathrm{A}}=\begin{bmatrix}\sqrt{\frac{\alpha_0\gamma_0}{\beta_0\delta_0}} & 0\\ 0 & e^{\imath\frac{\theta_0}{2}}\end{bmatrix},\quad \boldsymbol{M}_{\bar{\mathrm{A}}}=\begin{bmatrix}\sqrt{1-\frac{\alpha_0\gamma_0}{\beta_0\delta_0}} & 0\\ 0 & 0\end{bmatrix}, \tag{C3}$$

and Bob measures his qubit using local operators

$$\boldsymbol{M}_{\mathrm{B}}=\begin{bmatrix}e^{\imath\frac{\theta_0}{2}} & 0\\ 0 & \sqrt{\frac{\beta_0\gamma_0}{\alpha_0\delta_0}}\end{bmatrix},\quad \boldsymbol{M}_{\bar{\mathrm{B}}}=\begin{bmatrix}0 & 0\\ 0 & \sqrt{1-\frac{\beta_0\gamma_0}{\alpha_0\delta_0}}\end{bmatrix}. \tag{C4}$$

When the measurement results correspond to $\boldsymbol{M}_{\mathrm{A}}$ and $\boldsymbol{M}_{\mathrm{B}}$, the density matrix of the qubit pair after the measurement is given by

$$\breve{\boldsymbol{\rho}}=\frac{(\boldsymbol{M}_{\mathrm{A}}\otimes\boldsymbol{M}_{\mathrm{B}})\,\breve{\boldsymbol{\rho}}\,(\boldsymbol{M}_{\mathrm{A}}\otimes\boldsymbol{M}_{\mathrm{B}})^\dagger}{\mathrm{tr}\{(\boldsymbol{M}_{\mathrm{A}}\otimes\boldsymbol{M}_{\mathrm{B}})\,\breve{\boldsymbol{\rho}}\,(\boldsymbol{M}_{\mathrm{A}}\otimes\boldsymbol{M}_{\mathrm{B}})^\dagger\}}, \tag{C5}$$

where $\breve{\boldsymbol{\rho}}$ is the density matrix given in (6). Set the channel to be phase damping, i.e., $\eta=0$, then $\alpha=\beta=\gamma=\delta=\frac{1}{\sqrt{2}}$, and $\theta=0$. Further set the channel parameter p so that F equals to $\tilde{F}$ given by[6]

$$\tilde{F}=\frac{F_0}{F_0+(1-F_0)\frac{\gamma_0\delta_0}{\alpha_0\beta_0}}. \tag{C6}$$

Then according to (C5), the LOCC for RSSP transforms a density matrix $\breve{\boldsymbol{\rho}}$ with parameters $F=\tilde{F}$, $\alpha=\beta=\gamma=\delta=\frac{1}{\sqrt{2}}$, and $\theta=0$ to another density matrix given by

$$\begin{aligned}\breve{\boldsymbol{\rho}} &= F_0(\alpha_0|0\rangle+\beta_0|11\rangle)(\alpha_0\langle 00|+\beta_0 e\langle 11|)\\ &\quad+(1-F_0)(\gamma_0|01\rangle+\delta_0 e^{\imath\theta_0}|10\rangle)(\gamma_0\langle 01|+\delta_0 e^{-\imath\theta_0}\langle 10|),\end{aligned} \tag{C7}$$

the parameters of which are F_0, α_0, β_0, γ_0, δ_0, and θ_0. Let $\{\boldsymbol{N}_{\mathrm{A}}^{(k)},\boldsymbol{N}_{\mathrm{B}}^{(k)}\}_{k=1}^K$ be local operators that achieve the optimal fidelity $f(F_0,\alpha_0,\beta_0,\gamma_0,\delta_0,\theta_0)$ with initial density matrix $\breve{\boldsymbol{\rho}}$. Define new local operators

$$\boldsymbol{L}_{\mathrm{A}}^{(k)}=\boldsymbol{N}_{\mathrm{A}}^{(k)}\boldsymbol{M}_{\mathrm{A}},\quad \boldsymbol{L}_{\mathrm{B}}^{(k)}=\boldsymbol{N}_{\mathrm{B}}^{(k)}\boldsymbol{M}_{\mathrm{B}}.$$

Then $\{\boldsymbol{L}_{\mathrm{A}}^{(k)},\boldsymbol{L}_{\mathrm{B}}^{(k)}\}_{k=1}^K$ are valid local operators, and achieve the same fidelity $f(F_0,\alpha_0,\beta_0,\gamma_0,\delta_0,\theta_0)$ with initial density matrix $\breve{\boldsymbol{\rho}}$. Therefore, the optimal fidelity with initial density matrix $\breve{\boldsymbol{\rho}}$ is lower bounded by

$$f\left(\tilde{F},\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},0\right) \geqslant f(F_0,\alpha_0,\beta_0,\gamma_0,\delta_0,\theta_0). \tag{C8}$$

This together with (C2) gives

$$\begin{aligned}f\left(F,\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},0\right) &> \frac{F_0^2}{F_0^2+(1-F_0)^2\left(\frac{\gamma_0\delta_0}{\alpha_0\beta_0}\right)^2}\\ &= \frac{\tilde{F}^2}{\tilde{F}^2+(1-\tilde{F})^2}.\end{aligned} \tag{C9}$$

With (C9) the contradiction arises. This completes the proof.

APPENDIX D: PROOF OF LEMMA 3

Without loss of generality, denote the separable operator acting on two qubit pairs as $\boldsymbol{N}_{\mathrm{A}}\otimes\boldsymbol{N}_{\mathrm{B}}$, where $\boldsymbol{N}_{\mathrm{A}}$ and $\boldsymbol{N}_{\mathrm{B}}$ are employed by Alice and Bob, respectively. Every operator $\boldsymbol{N}$ for two qubits can be written equivalently as $\boldsymbol{N}=\boldsymbol{N}(\boldsymbol{H}^\dagger\otimes\boldsymbol{H}^\dagger)(\boldsymbol{H}\otimes\boldsymbol{H})$, where $\boldsymbol{H}=\frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right]$ is the Hadamard operator. Therefore, denote $\tilde{\boldsymbol{N}}_{\mathrm{X}}=\boldsymbol{N}_{\mathrm{X}}(\boldsymbol{H}^\dagger\otimes\boldsymbol{H}^\dagger)$, $X\in\{A,B\}$, then the separable operator $\boldsymbol{N}_{\mathrm{A}}\otimes\boldsymbol{N}_{\mathrm{B}}$ for two qubit pairs is equivalent to first perform $\boldsymbol{H}$ on every qubit, and then perform $\tilde{\boldsymbol{N}}_{\mathrm{A}}\otimes\tilde{\boldsymbol{N}}_{\mathrm{B}}$.

From (6), $\breve{\boldsymbol{\rho}}=F|\Phi^+\rangle\langle\Phi^+|+(1-F)|\Psi^+\rangle\langle\Psi^+|$ when the channel is phase damping. Hence after performing Hadamard operation on the qubits, the density matrix of the qubit pair becomes

$$\begin{aligned}\tilde{\boldsymbol{\rho}} &= (\boldsymbol{H}\otimes\boldsymbol{H})\,\breve{\boldsymbol{\rho}}\,(\boldsymbol{H}\otimes\boldsymbol{H})^\dagger\\ &= F|\Phi^+\rangle\langle\Phi^+|+(1-F)|\Phi^-\rangle\langle\Phi^-|.\end{aligned} \tag{D1}$$

Therefore, the joint density matrix of two qubit pairs, where the first and last two qubits belong to Alice Bob, respectively, is given by

$$\begin{aligned}\boldsymbol{\rho}_{\mathrm{J}} &= \boldsymbol{P}\,(\tilde{\boldsymbol{\rho}}\otimes\tilde{\boldsymbol{\rho}})\,\boldsymbol{P}\\ &= F^2|\Phi^{(1)}\rangle\langle\Phi^{(1)}|+F(1-F)(|\Phi^{(2)}\rangle\langle\Phi^{(2)}|+|\Phi^{(3)}\rangle\langle\Phi^{(3)}|)\\ &\quad+(1-F)^2|\Phi^{(4)}\rangle\langle\Phi^{(4)}|,\end{aligned} \tag{D2}$$

where $\boldsymbol{P}$ is the permutation operator that switches the second and third qubits:

$$\begin{aligned}|\Phi^{(1)}\rangle &= \tfrac{1}{2}(|0000\rangle+|0101\rangle+|1010\rangle+|1111\rangle),\\ |\Phi^{(2)}\rangle &= \tfrac{1}{2}(|0000\rangle-|0101\rangle+|1010\rangle-|1111\rangle),\\ |\Phi^{(3)}\rangle &= \tfrac{1}{2}(|0000\rangle+|0101\rangle-|1010\rangle-|1111\rangle),\\ |\Phi^{(4)}\rangle &= \tfrac{1}{2}(|0000\rangle-|0101\rangle-|1010\rangle+|1111\rangle).\end{aligned} \tag{D3}$$

From (D2), after operator $\tilde{\boldsymbol{N}}_{\mathrm{A}}\otimes\tilde{\boldsymbol{N}}_{B}$ acts on the two qubit pairs, the density matrix of the first qubit pair is given

[6]Equations (10) and (11) imply $0\leqslant\gamma\leqslant\beta\leqslant\frac{1}{\sqrt{2}}\leqslant\alpha\leqslant\delta\leqslant 1$ and $\alpha^2+\beta^2=\gamma^2+\delta^2=1$, showing that $\frac{\gamma\delta}{\alpha\beta}\in[0,1]$ for all valid α,β,γ, and δ. Hence, $\tilde{F}$ in (C6) is in the interval $(\frac{1}{2},1]$ as long as $F_0\in(\frac{1}{2},1]$. This guarantees the existence of p.

by

$$\breve{\boldsymbol{\rho}} = \frac{\mathrm{tr}_{2,4}\left\{\sum_{i=1}^{4} C_i(\tilde{\boldsymbol{N}}_{\mathrm{A}} \otimes \tilde{\boldsymbol{N}}_B)\,|\Phi^{(i)}\rangle\langle\Phi^{(i)}|\,(\tilde{\boldsymbol{N}}_{\mathrm{A}} \otimes \tilde{\boldsymbol{N}}_B)^\dagger\right\}}{\mathrm{tr}\left\{\sum_{i=1}^{4} C_i(\tilde{\boldsymbol{N}}_{\mathrm{A}} \otimes \tilde{\boldsymbol{N}}_B)\,|\Phi^{(i)}\rangle\langle\Phi^{(i)}|\,(\tilde{\boldsymbol{N}}_{\mathrm{A}} \otimes \tilde{\boldsymbol{N}}_B)^\dagger\right\}} = \frac{\sum_{i=1}^{4} C_i \mathrm{tr}_{2,4}\{\boldsymbol{\phi}^{(i)}\boldsymbol{\phi}^{(i)\dagger}\}}{\mathrm{tr}\left\{\sum_{i=1}^{4} C_i \mathrm{tr}_{2,4}\{\boldsymbol{\phi}^{(i)\dagger}\boldsymbol{\phi}^{(i)}\}\right\}}, \tag{D4}$$

where $C_1 = F^2$, $C_2 = C_3 = F(1-F)$, $C_4 = (1-F)^2$, and $\boldsymbol{\phi}^{(i)} = (\tilde{\boldsymbol{N}}_{\mathrm{A}} \otimes \tilde{\boldsymbol{N}}_B)|\Phi^{(i)}\rangle$. Denote $|w\rangle = |00\rangle$, $|x\rangle = |01\rangle$, $|y\rangle = |10\rangle$, and $|z\rangle = |11\rangle$, and denote

$$\boldsymbol{\psi}^{(i)} = \mathrm{tr}_{2,4}\{\boldsymbol{\phi}^{(i)}\}.$$

Then

$$\begin{aligned}\boldsymbol{\psi}^{(i)} &= \mathrm{tr}_{2,4}\{(\tilde{\boldsymbol{N}}_{\mathrm{A}} \otimes \tilde{\boldsymbol{N}}_B)\,|\Phi^{(i)}\rangle\} \\ &= \left(\sum_{k=0}^{1}(\mathbb{I}_2 \otimes \langle k|)\tilde{\boldsymbol{N}}_{\mathrm{A}} \otimes \sum_{j=0}^{1}(\mathbb{I}_2 \otimes \langle j|)\tilde{\boldsymbol{N}}_B\right) \\ &\quad \times \begin{bmatrix}|ww\rangle & |xx\rangle & |yy\rangle & |zz\rangle\end{bmatrix}\boldsymbol{v}^{(i)} \\ &= [\boldsymbol{w} \quad \boldsymbol{x} \quad \boldsymbol{y} \quad \boldsymbol{z}]\boldsymbol{v}^{(i)},\end{aligned} \tag{D5}$$

where $\boldsymbol{v}^{(i)}$ is the ith column of the unitary matrix $\boldsymbol{V}$ defined in (17) and

$$\begin{aligned}\boldsymbol{s} &= \left(\sum_{k=0}^{1}(\mathbb{I}_2 \otimes \langle k|)\tilde{\boldsymbol{N}}_{\mathrm{A}} \otimes \sum_{j=0}^{1}(\mathbb{I}_2 \otimes \langle j|)\tilde{\boldsymbol{N}}_B\right)|ss\rangle \\ &= \begin{bmatrix}s_{11}\\ s_{12}\\ s_{21}\\ s_{22}\end{bmatrix},\end{aligned} \tag{D6}$$

where $s \in \{w,x,y,z\}$. Combining (D4) and (D5) gives (14) and (15).

In (D6), $|ss\rangle$ is a separable state, and

$$\sum_{k=0}^{1}(\mathbb{I}_2 \otimes \langle k|)\tilde{\boldsymbol{N}}_{\mathrm{A}} \otimes \sum_{j=0}^{1}(\mathbb{I}_2 \otimes \langle j|)\tilde{\boldsymbol{N}}_B$$

is a separable operator. Therefore, vectors $\boldsymbol{s}$, $s \in \{w,x,y,z\}$ must also be separable. As 1×4 vectors, $\boldsymbol{s}$ are separable if and only if

$$s_{11}s_{22} = s_{12}s_{21}, \quad \forall s \in \{w,x,y,z\},$$

which give (16).

Finally, the probability that $\boldsymbol{N}_{\mathrm{A}} \otimes \boldsymbol{N}_{\mathrm{B}}$ acts on the qubits must not be zero, which implies

$$\sum_{i=1}^{4} C_i \boldsymbol{\psi}^{(i)\dagger}\boldsymbol{\psi}^{(i)} > 0.$$

This completes the proof.

APPENDIX E: PROOF OF THEOREM 2

We will first prove that the fidelity F^* given in (18) is an upper bound. From Lemma 2, it is sufficient to prove the upper bound for the special case when the channel is phase damping.

Express the LOCC performed by the agents as $\boldsymbol{N}_{\mathrm{A}}^{(k)} \otimes \boldsymbol{N}_{\mathrm{B}}^{(k)}$, $k \in \{1,2,\ldots,K\}$. Without loss of generality, assume $\boldsymbol{N}_{\mathrm{A}}^{(1)} \otimes \boldsymbol{N}_{\mathrm{B}}^{(1)}$ is one of the operators that lead to the highest fidelity. Then from Lemma 3, conditioned on the event that $\boldsymbol{N}_{\mathrm{A}}^{(1)} \otimes \boldsymbol{N}_{\mathrm{B}}^{(1)}$ acts on the two qubit pairs, the fidelity of the kept qubit pair is given by

$$\begin{aligned}&\langle\Phi^+|\breve{\boldsymbol{\rho}}|\Phi^+\rangle \\ &= \frac{\sum_{i=1}^{4} C_i\langle\Phi^+|\,\boldsymbol{\psi}^{(i)}\boldsymbol{\psi}^{(i)\dagger}|\Phi^+\rangle}{\sum_{i=1}^{4} C_i \boldsymbol{\psi}^{(i)\dagger}\boldsymbol{\psi}^{(i)}} \\ &= \frac{\frac{1}{2}\sum_{i=1}^{4} C_i\left|\sum_{k=1}^{2}[w_{kk} \quad x_{kk} \quad y_{kk} \quad z_{kk}]\boldsymbol{v}^{(i)}\right|^2}{\sum_{i=1}^{4} C_i \sum_{k=1}^{2}\sum_{j=1}^{2}|[w_{kj} \quad x_{kj} \quad y_{kj} \quad z_{kj}]\boldsymbol{v}^{(i)}|^2},\end{aligned}$$

where C_i, $\boldsymbol{v}^{(i)}$, $i \in \{1,2,3,4\}$ and s_{kj}, $s \in \{w,x,y,z\}$, $k,j \in \{1,2\}$ are defined in Lemma 3. Note that from Theorem 1, $F \in [\frac{1}{2},1]$, implying that $C_1 \geqslant C_2 = C_3 \geqslant C_4 \geqslant 0$. Therefore, to prove the upper bound part of Theorem 2, it is sufficient to show that the following proposition is true.

Proposition 1. Maximum fidelity. For any $s_{kj} \in \mathbb{C}$, $s \in \{w,x,y,z\}$, $k,j \in \{1,2\}$, and $1 \geqslant C_1 \geqslant C_2 \geqslant C_3 \geqslant C_4 \geqslant 0$, satisfying $s_{11}s_{22} = s_{12}s_{21}$ and

$$\sum_{i=1}^{4} C_i \sum_{k=1}^{2}\sum_{j=1}^{2}|[w_{kj} \quad x_{kj} \quad y_{kj} \quad z_{kj}]\boldsymbol{v}^{(i)}|^2 > 0,$$

the following inequality holds:

$$\frac{\sum_{i=1}^{4} C_i\left|\sum_{k=1}^{2}[w_{kk} \quad x_{kk} \quad y_{kk} \quad z_{kk}]\boldsymbol{v}^{(i)}\right|^2}{\sum_{i=1}^{4} C_i \sum_{k=1}^{2}\sum_{j=1}^{2}|[w_{kj} \quad x_{kj} \quad y_{kj} \quad z_{kj}]\boldsymbol{v}^{(i)}|^2} \leqslant \frac{2C_1}{C_1 + C_4}. \tag{E1}$$

To prove Proposition 1, first simplify (E1) via the following lemma.

Lemma 4. Simplify parameters. Consider coefficients $r_1,r_2,\breve{r}_2,r_3,\breve{r}_3,r_4 \geqslant 0$, $\breve{r}_4 > 0$ and variable $t \in [0,1]$, satisfying $r_3t + r_4 > 0$ and

$$r_2\breve{r}_4 - \breve{r}_2 r_4 \leqslant 0. \tag{E2}$$

If inequality

$$\frac{r_1t + r_2}{r_3t + r_4} \leqslant \frac{\breve{r}_2}{\breve{r}_3t + \breve{r}_4} \tag{E3}$$

holds for $t = \breve{t} \geqslant 0$, then it holds for all $t \in [0,\breve{t}]$.

Proof. Define function

$$f(t) \triangleq r_1\breve{r}_3t^2 + (r_2\breve{r}_3 + r_1\breve{r}_4 - \breve{r}_2r_3)t + r_2\breve{r}_4 - \breve{r}_2r_4.$$

From (E2), $f(0) \leqslant 0$. Since $r_3t + r_4 > 0$ and $\breve{r}_3t + \breve{r}_4 > 0$, the fact that (E3) holds for $t = \breve{t}$ is equivalent to $f(\breve{C}) \leqslant 0$. Moreover, since $f''(t) = r_1\breve{r}_3 \geqslant 0$, $f(t)$ is a convex function. Therefore, $f(t) \leqslant 0$, $\forall t \in [0,\breve{t}]$, which is equivalent to (E3) holds $\forall t \in [0,\breve{t}]$. This completes the proof of Lemma 4. □

Letting $C_4 = t$,

$$\left|\sum_{k=1}^{2}[w_{kk}\quad x_{kk}\quad y_{kk}\quad z_{kk}]\boldsymbol{v}^{(4)}\right|^2 = r_1,$$

$$\sum_{i=1}^{3} C_i\left|\sum_{k=1}^{2}[w_{kk}\quad x_{kk}\quad y_{kk}\quad z_{kk}]\boldsymbol{v}^{(i)}\right|^2 = r_2,$$

$$\sum_{k=1}^{2}\sum_{j=1}^{2}|[w_{kj}\quad x_{kj}\quad y_{kj}\quad z_{kj}]\boldsymbol{v}^{(4)}|^2 = r_3,$$

$$\sum_{i=1}^{3} C_i\sum_{k=1}^{2}\sum_{j=1}^{2}|[w_{kj}\quad x_{kj}\quad y_{kj}\quad z_{kj}]\boldsymbol{v}^{(i)}|^2 = r_4.$$

$2C_1 = \check{r}_2$, $1 = \check{r}_3$, and $C_1 = \check{r}_4$ in (E1) gives the form of (E3). It can be verified that

$$\begin{aligned}
&r_2\check{r}_4 - \check{r}_2 r_4\\
&= C_1\sum_{i=1}^{3} C_i\left(\left|\sum_{k=1}^{2}[w_{kk}\quad x_{kk}\quad y_{kk}\quad z_{kk}]\boldsymbol{v}^{(i)}\right|^2\right.\\
&\quad\left. - 2\sum_{k=1}^{2}\sum_{j=1}^{2}|[w_{kj}\quad x_{kj}\quad y_{kj}\quad z_{kj}]\boldsymbol{v}^{(i)}|^2\right)\\
&\leqslant 2C_1\sum_{i=1}^{3} C_i\left(\sum_{k=1}^{2}|[w_{kk}\quad x_{kk}\quad y_{kk}\quad z_{kk}]\boldsymbol{v}^{(i)}|^2\right.\\
&\quad\left. - \sum_{k=1}^{2}|[w_{kk}\quad x_{kk}\quad y_{kk}\quad z_{kk}]\boldsymbol{v}^{(i)}|^2\right)\\
&= 0.
\end{aligned}$$

Therefore, Lemma 4 shows that (E3) is true $\forall t \in [0, C_3]$ if it is true for $t = C_3$. This implies that to prove Proposition 1, it is sufficient to prove (E1) for the case of $C_4 = C_3$. Repeating this process two more times, i.e., applying Lemma 4 to (E1) with $C_3 = t$, and then with $C_2 = t$, it can be shown that considering the case in which $C_1 = C_2 = C_3 = C_4$ is sufficient to prove the proposition. Then, (E1) simplifies to

$$\frac{\sum_{i=1}^{4}\left|\sum_{k=1}^{2}[w_{kk}\quad x_{kk}\quad y_{kk}\quad z_{kk}]\boldsymbol{v}^{(i)}\right|^2}{\sum_{k=1}^{2}\sum_{j=1}^{2}\sum_{i=1}^{4}|[w_{kj}\quad x_{kj}\quad y_{kj}\quad z_{kj}]\boldsymbol{v}^{(i)}|^2} \leqslant 1. \quad \text{(E4)}$$

Note that

$$\begin{aligned}
&\sum_{i=1}^{4}|[w_{kj}\quad x_{kj}\quad y_{kj}\quad z_{kj}]\boldsymbol{v}^{(i)}|^2\\
&= \sum_{i=1}^{4}[w_{kj}\quad x_{kj}\quad y_{kj}\quad z_{kj}]\boldsymbol{v}^{(i)}\boldsymbol{v}^{(i)\dagger}[w_{kj}\quad x_{kj}\quad y_{kj}\quad z_{kj}]^\dagger\\
&= [w_{kj}\quad x_{kj}\quad y_{kj}\quad z_{kj}]\boldsymbol{V}\boldsymbol{V}^\dagger[w_{kj}\quad x_{kj}\quad y_{kj}\quad z_{kj}]^\dagger\\
&= \sum_{s\in\{w,x,y,z\}}|s_{kj}|^2, \quad \text{(E5)}
\end{aligned}$$

where the last equality is due to the fact that $\boldsymbol{V}$ is unitary. Similarly,

$$\sum_{i=1}^{4}\left|\sum_{k=1}^{2}[w_{kk}\quad x_{kk}\quad y_{kk}\quad z_{kk}]\boldsymbol{v}^{(i)}\right|^2 = \sum_{s\in\{w,x,y,z\}}|s_{11}+s_{22}|^2. \quad \text{(E6)}$$

Then it can be obtained that

$$\begin{aligned}
&\sum_{k=1}^{2}\sum_{j=1}^{2}\sum_{i=1}^{4}|[w_{kj}\quad x_{kj}\quad y_{kj}\quad z_{kj}]\boldsymbol{v}^{(i)}|^2\\
&\overset{(a)}{=} \sum_{s\in\{w,x,y,z\}}(|s_{11}|^2+|s_{22}|^2) + \sum_{s\in\{w,x,y,z\}}(|s_{12}|^2+|s_{21}|^2)\\
&\overset{(b)}{\geqslant} \sum_{s\in\{w,x,y,z\}}(|s_{11}|^2+|s_{22}|^2) + 2\sum_{s\in\{w,x,y,z\}}|s_{11}||s_{22}|\\
&\overset{(c)}{=} \sum_{i=1}^{4}\left|\sum_{k=1}^{2}[w_{kk}\quad x_{kk}\quad y_{kk}\quad z_{kk}]\boldsymbol{v}^{(i)}\right|^2, \quad \text{(E7)}
\end{aligned}$$

where (a), (b), and (c) are due to (E5), (16), and (E6), respectively. This inequality shows that (E4) is true, which then proves Proposition 1. This proves that the fidelity F^* given in (18) is an upper bound.

Finally, we use the constructive method to show that fidelity F^* given in (18) is achievable. In fact, (22) of Theorem 3 shows that the fidelity in (18) is achieved by adopting the RSSP and the first round distillation of the algorithm proposed in Sec. III C and keeping a qubit pair only if measurement results correspond to $|1\rangle\langle 1|$. This completes the proof.

APPENDIX F: PROOF OF THEOREM 3

First, the following lemma summarizes the effect of RSSP.

Lemma 5. Performance of RSSP. In process of RSSP, qubit pairs are kept with probability

$$P_s = 2F_0\beta^2 + (1-F_0)\left(\gamma^2 + \frac{\beta^2\delta^2}{\alpha^2}\right). \quad \text{(F1)}$$

For a kept qubit pair, its density matrix is given by

$$\tilde{\boldsymbol{\rho}} = \tilde{F}|\Phi^+\rangle\langle\Phi^+| + (1-\tilde{F})|\tilde{\nu}\rangle\langle\tilde{\nu}|, \quad \text{(F2)}$$

where

$$|\tilde{\nu}\rangle = \tilde{\gamma}|01\rangle + \tilde{\delta}e^{\iota\theta}|10\rangle \quad \text{(F3)}$$

with

$$\tilde{F} = \frac{2F_0\alpha^2\beta^2}{2F_0\alpha^2\beta^2 + (1-F_0)(\alpha^2\gamma^2+\beta^2\delta^2)}, \quad \text{(F4)}$$

$$\tilde{\gamma} = \frac{\alpha\gamma}{\sqrt{\alpha^2\gamma^2+\beta^2\delta^2}}, \quad \text{(F5)}$$

$$\tilde{\delta} = \frac{\beta\delta}{\sqrt{\alpha^2\gamma^2+\beta^2\delta^2}}. \quad \text{(F6)}$$

Proof. The qubit pairs are kept with probability

$$\mathrm{tr}\{(\mathbb{I}_2\otimes\boldsymbol{M}_B)\check{\boldsymbol{\rho}}(\mathbb{I}_2\otimes\boldsymbol{M}_B)^\dagger\} \quad \text{(F7)}$$

and the density matrix of a kept qubit pair is given by

$$\tilde{\boldsymbol{\rho}} = \frac{(\mathbb{I}_2 \otimes \boldsymbol{M}_\mathrm{B})\, \check{\boldsymbol{\rho}}\, (\mathbb{I}_2 \otimes \boldsymbol{M}_\mathrm{B})^\dagger}{\mathrm{tr}\{(\mathbb{I}_2 \otimes \boldsymbol{M}_\mathrm{B})\, \check{\boldsymbol{\rho}}\, (\mathbb{I}_2 \otimes \boldsymbol{M}_\mathrm{B})^\dagger\}}. \tag{F8}$$

Substituting (6)–(11) into (F7) and (F8), one can obtain (F2)–(F6). The details are omitted for brevity. □

From (F2) and (F3), after the RSSP, the joint density matrix of two qubit pairs, where the first and last two qubits belong to Alice and Bob, respectively, is given by

$$\begin{aligned}\boldsymbol{\rho}_\mathrm{J} &= \boldsymbol{P}\tilde{\boldsymbol{\rho}} \otimes \tilde{\boldsymbol{\rho}}\, \boldsymbol{P}^\dagger \\ &= \tilde{F}^2|\Omega^{(1)}\rangle\langle\Omega^{(1)}| + \tilde{F}(1-\tilde{F})(|\Omega^{(2)}\rangle\langle\Omega^{(2)}| + |\Omega^{(3)}\rangle\langle\Omega^{(3)}|) \\ &\quad + (1-\tilde{F})^2|\Omega^{(4)}\rangle\langle\Omega^{(4)}|,\end{aligned}$$

where $\boldsymbol{P}$ is the permutation operator that switches the second and third qubits, and

$$\begin{aligned}|\Omega^{(1)}\rangle &= \frac{1}{2}|0000\rangle + \frac{1}{2}|0101\rangle + \frac{1}{2}|1010\rangle + \frac{1}{2}|1111\rangle, \\ |\Omega^{(2)}\rangle &= \frac{\tilde{\gamma}\sqrt{2}}{2}|0001\rangle + \frac{\tilde{\delta}e^{\imath\theta}\sqrt{2}}{2}|0100\rangle \\ &\quad + \frac{\tilde{\gamma}\sqrt{2}}{2}|1011\rangle + \frac{\tilde{\delta}e^{\imath\theta}\sqrt{2}}{2}|1110\rangle, \\ |\Omega^{(3)}\rangle &= \frac{\tilde{\gamma}\sqrt{2}}{2}|0010\rangle + \frac{\tilde{\gamma}\sqrt{2}}{2}|0111\rangle \\ &\quad + \frac{\tilde{\delta}e^{\imath\theta}\sqrt{2}}{2}|1000\rangle + \frac{\tilde{\delta}e^{\imath\theta}\sqrt{2}}{2}|1101\rangle, \\ |\Omega^{(4)}\rangle &= \tilde{\gamma}^2|0011\rangle + \tilde{\gamma}\tilde{\delta}e^{\imath\theta}|0110\rangle \\ &\quad + \tilde{\gamma}\tilde{\delta}e^{\imath\theta}|1001\rangle + \tilde{\delta}^2e^{\imath 2\theta}|1100\rangle.\end{aligned}$$

For the first round of distillation, after both agents perform the CNOT operation, the joint density matrix of two qubit pairs becomes

$$\begin{aligned}\check{\boldsymbol{\rho}}_\mathrm{J} &= \tilde{F}^2|\check{\Omega}^{(1)}\rangle\langle\check{\Omega}^{(1)}| + \tilde{F}(1-\tilde{F})(|\check{\Omega}^{(2)}\rangle\langle\check{\Omega}^{(2)}| + |\check{\Omega}^{(3)}\rangle\langle\check{\Omega}^{(3)}|) \\ &\quad + (1-\tilde{F})^2|\check{\Omega}^{(4)}\rangle\langle\check{\Omega}^{(4)}|,\end{aligned} \tag{F9}$$

where

$$\begin{aligned}|\check{\Omega}^{(1)}\rangle &= \frac{1}{2}|0000\rangle + \frac{1}{2}|0101\rangle + \frac{1}{2}|1111\rangle + \frac{1}{2}|1010\rangle, \\ |\check{\Omega}^{(2)}\rangle &= \frac{\tilde{\gamma}\sqrt{2}}{2}|0001\rangle + \frac{\tilde{\delta}e^{\imath\theta}\sqrt{2}}{2}|0100\rangle \\ &\quad + \frac{\tilde{\gamma}\sqrt{2}}{2}|1110\rangle + \frac{\tilde{\delta}e^{\imath\theta}\sqrt{2}}{2}|1011\rangle, \\ |\check{\Omega}^{(3)}\rangle &= \frac{\tilde{\gamma}\sqrt{2}}{2}|0011\rangle + \frac{\tilde{\gamma}\sqrt{2}}{2}|0110\rangle \\ &\quad + \frac{\tilde{\delta}e^{\imath\theta}\sqrt{2}}{2}|1100\rangle + \frac{\tilde{\delta}e^{\imath\theta}\sqrt{2}}{2}|1001\rangle, \\ |\check{\Omega}^{(4)}\rangle &= \tilde{\gamma}^2|0010\rangle + \tilde{\gamma}\tilde{\delta}e^{\imath\theta}|0111\rangle \\ &\quad + \tilde{\gamma}\tilde{\delta}e^{\imath\theta}|1101\rangle + \tilde{\delta}^2e^{\imath 2\theta}|1000\rangle.\end{aligned}$$

From (F9), if both measurement results correspond to $|1\rangle\langle 1|$, the (un-normalized) density matrix of the source qubit pair is given by

$$\begin{aligned}\boldsymbol{\rho}_{11} &= (\mathbb{I}_2 \otimes \langle 1| \otimes \mathbb{I}_2 \otimes \langle 1|)\, \check{\boldsymbol{\rho}}_\mathrm{J}\, (\mathbb{I}_2 \otimes |1\rangle \otimes \mathbb{I}_2 \otimes |1\rangle) \\ &= \tilde{F}^2\tfrac{1}{2}|\Phi^+\rangle\langle\Phi^+| \\ &\quad + (1-\tilde{F})^2(\tilde{\gamma}\tilde{\delta})^2(|01\rangle + |10\rangle)(\langle 01| + \langle 10|).\end{aligned} \tag{F10}$$

Otherwise, if both measurement results correspond to $|0\rangle\langle 0|$, the (un-normalized) density matrix of the source qubit pair is given by

$$\begin{aligned}\boldsymbol{\rho}_{00} &= (\mathbb{I}_2 \otimes \langle 0| \otimes \mathbb{I}_2 \otimes \langle 0|)\, \check{\boldsymbol{\rho}}_\mathrm{J}\, (\mathbb{I}_2 \otimes |0\rangle \otimes \mathbb{I}_2 \otimes |0\rangle) \\ &= \tilde{F}^2\tfrac{1}{2}|\Phi^+\rangle\langle\Phi^+| + (1-\tilde{F})^2(\tilde{\gamma}^2|01\rangle + \tilde{\delta}^2e^{\imath 2\theta}|10\rangle) \\ &\quad \times (\tilde{\gamma}^2\langle 01| + \tilde{\delta}^2e^{-\imath 2\theta}\langle 10|).\end{aligned} \tag{F11}$$

From (F10) and (F11), if the agents adopt the FP approach, i.e., keep the source qubit pair only if both measurement results correspond to $|1\rangle\langle 1|$, the probability of keeping the source qubit pair is

$$P_\mathrm{f} = \mathrm{tr}\{\boldsymbol{\rho}_{11}\} = \frac{\tilde{F}^2}{2} + 2(1-\tilde{F})^2(\tilde{\gamma}\tilde{\delta})^2, \tag{F12}$$

the fidelity of the kept qubit pairs is

$$F_1 = \frac{\frac{1}{2}\tilde{F}^2}{P_f} = \frac{\tilde{F}^2}{\tilde{F}^2 + 4(1-\tilde{F})^2(\tilde{\gamma}\tilde{\delta})^2}, \tag{F13}$$

and the density matrix of the kept qubit pair is

$$\tilde{\boldsymbol{\rho}} = \frac{\boldsymbol{\rho}_{11}}{P_f} = F_1|\Phi^+\rangle\langle\Phi^+| + (1-F_1)|\Psi^+\rangle\langle\Psi^+|. \tag{F14}$$

If the agents adopt the PP approach, i.e., keeping the source qubit pair if the measurement results match, the probability of preserving the source qubit pair is

$$P_\mathrm{p} = \mathrm{tr}\{\boldsymbol{\rho}_{11} + \boldsymbol{\rho}_{00}\} = \tilde{F}^2 + (1-\tilde{F})^2, \tag{F15}$$

the fidelity of the kept qubit pairs is

$$F_1 = \frac{\frac{1}{2}\tilde{F}^2 + \frac{1}{2}\tilde{F}^2}{P_1} = \frac{\tilde{F}^2}{\tilde{F}^2 + (1-\tilde{F})^2}, \tag{F16}$$

and the density matrix of the kept qubit pair can be written as

$$\tilde{\boldsymbol{\rho}} = \frac{\boldsymbol{\rho}_{11} + \boldsymbol{\rho}_{00}}{P_\mathrm{p}} = F_1|\Phi^+\rangle\langle\Phi^+| + G|\Psi\rangle\langle\Psi| + \tilde{G}|\tilde{\Psi}\rangle\langle\tilde{\Psi}|, \tag{F17}$$

where $G + \tilde{G} = 1 - F_1$, $|\Psi\rangle, |\tilde{\Psi}\rangle \in \mathrm{span}(|01\rangle, |10\rangle)$, and $\langle\Psi|\tilde{\Psi}\rangle = 0$.

From Lemma 5, (F12), (F13), (F15), and (F16), after the RSSP and the first round of distillation, a qubit pair is kept with

probability

$$P_1 = \begin{cases} \frac{P_s P_f}{2} = \frac{F_0^2\alpha^2\beta^4+(1-F_0)^2\beta^2\gamma^2\delta^2}{2F_0\alpha^2\beta^2+(1-F_0)(\alpha^2\gamma^2+\beta^2\delta^2)} & \text{for the FP approach} \\ \frac{P_s P_p}{2} = \frac{4F_0^2\alpha^4\beta^4+(1-F_0)^2(\alpha^2\gamma^2+\beta^2\delta^2)^2}{4F_0\alpha^4\beta^2+2(1-F_0)\alpha^2(\alpha^2\gamma^2+\beta^2\delta^2)} & \text{for the PP approach} \end{cases}$$

and fidelity

$$F_1 = \begin{cases} \frac{F_0^2}{F_0^2+(1-F_0)^2\left(\frac{\gamma\delta}{\alpha\beta}\right)^2} & \text{for the FP approach} \\ \frac{F_0^2}{F_0^2+\frac{1}{4}(1-F_0)^2\left(\frac{\gamma^2}{\beta^2}+\frac{\delta^2}{\alpha^2}\right)^2} & \text{for the PP approach} \end{cases}.$$

For the following rounds of distillations, one can take (F14) or (F17) as input and use similar analysis as in (F9)–(F11) and (F15)–(F17). This analysis will show that

$$P_k = \frac{1}{2}\left[F_{k-1}^2 + (1-F_{k-1})^2\right], \quad F_k = \frac{F_{k-1}^2}{F_{k-1}^2 + (1-F_{k-1})^2}$$

and the density matrix of the kept qubit pairs maintains the same structure as in (F14) or (F17). This competes the proof.

[1] S. Lloyd, Phys. Rev. Lett. **90**, 167902 (2003).
[2] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).
[3] A. E. Ulanov, I. A. Fedorov, A. A. Pushkina, Y. V. Kurochkin, T. C. Ralph, and A. I. Lvovsky, Nat. Photonics **9**, 764 (2015).
[4] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).
[5] M. Koashi and J. Preskill, Phys. Rev. Lett. **90**, 057902 (2003).
[6] D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill, in *Proceedings of the IEEE International Symposium on Information Theory* (IEEE, Piscataway, NJ, 2004), p. 136.
[7] C. H. Bennett and S. J. Wiesner, Phys. Rev. Lett. **69**, 2881 (1992).
[8] C. Wang, F.-G. Deng, Y.-S. Li, X.-S. Liu, and G. L. Long, Phys. Rev. A **71**, 044305 (2005).
[9] J. T. Barreiro, T.-C. Wei, and P. G. Kwiat, Nat. Phys. **4**, 282 (2008).
[10] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).
[11] M. A. Nielsen, E. Knill, and R. Laflamme, Nature (London) **396**, 52 (1998).
[12] D. Gottesman and I. L. Chuang, Nature (London) **402**, 390 (1999).
[13] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, Phys. Rev. Lett. **83**, 3081 (1999).
[14] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, IEEE Trans. Inf. Theory **48**, 2637 (2002).
[15] P. W. Shor, in *Quantum Information, Statistics, Probability: Dedicated to Alexander S. Holevo on the Occasion of his 60th Birthday*, edited by O. Hirota (Rinton, Princeton, NJ, 2004), pp. 144–152.
[16] A. S. Holevo, Probl. Inform. Transmission **48**, 3 (2012).
[17] A. S. Holevo and M. E. Shirokov, Mathematical Notes **97**, 974 (2015).
[18] N. Sangouard, C. Simon, H. de Riedmatten, and N. Gisin, Rev. Mod. Phys. **83**, 33 (2011).
[19] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, Phys. Rev. Lett. **76**, 722 (1996).
[20] D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera, Phys. Rev. Lett. **77**, 2818 (1996).
[21] T. Opatrný and G. Kurizki, Phys. Rev. A **60**, 167 (1999).
[22] H. F. Chau and K. H. Ho, Quant. Info. Proc. **10**, 213 (2011).
[23] J. Dehaene, M. Van den Nest, B. De Moor, and F. Verstraete, Phys. Rev. A **67**, 022310 (2003).
[24] Karl Gerd H. Vollbrecht and F. Verstraete, Phys. Rev. A **71**, 062325 (2005).
[25] E. Hostens, J. Dehaene, and B. De Moor, Phys. Rev. A **73**, 062337 (2006).
[26] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).
[27] R. Matsumoto, J. Phys. A **36**, 8113 (2003).
[28] A. Ambainis and D. Gottesman, IEEE Trans. Inf. Theory **52**, 748 (2006).
[29] S. Watanabe, R. Matsumoto, and T. Uyem, J. Phys. A **39**, 4273 (2006).
[30] N. H. Nickerson, J. F. Fitzsimons, and S. C. Benjamin, Phys. Rev. X **4**, 041041 (2014).
[31] P. Kómár, E. M. Kessler, M. Bishof, L. Jiang, A. S. Sørensen, J. Ye, and M. D. Lukin, Nature Physics **10**, 582 (2014).
[32] N. Kalb, A. A. Reiserer, P. C. Humphreys, J. J. W. Bakermans, S. J. Kamerling, N. H. Nickerson, S. C. Benjamin, D. J. Twitchen, M. Markham, and R. Hanson, Science **356**, 928 (2017).
[33] P. W. Shor, Phys. Rev. A **52**, R2493(R) (1995).
[34] D. Gottesman, Phys. Rev. A **54**, 1862 (1996).
[35] E. Knill, Nature (London) **434**, 39 (2005).
[36] D. Gottesman, in *Quantum Information Science and Its Contributions to Mathematics* (American Mathematical Society, Providence, 2009), pp. 13–58.
[37] W. G. Unruh, Phys. Rev. A **51**, 992 (1995).
[38] D. P. Divincenzo, Fortschr. Phys. **48**, 771 (2000).
[39] K.-A. Suominen, *Handbook of Natural Computing* (Springer, Berlin, Heidelberg, 2012), pp. 1493–1520.
[40] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O'Brien, Nature (London) **464**, 45 (2010).
[41] A. S. Fletcher, P. W. Shor, and M. Z. Win, Phys. Rev. A **75**, 012338 (2007).
[42] A. S. Fletcher, P. W. Shor, and M. Z. Win, Phys. Rev. A **77**, 012320 (2008).
[43] A. S. Fletcher, P. W. Shor, and M. Z. Win, IEEE Trans. Inf. Theory **54**, 5705 (2008).

[44] J. Preskill, Lecture notes for Physics 219 (2015).
[45] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University, Cambridge, England, 2000).
[46] C. H. Bennett, D. P. DiVincenzo, P. W. Shor, J. A. Smolin, B. M. Terhal, and W. K. Wootters, Phys. Rev. Lett. **87**, 077902 (2001).
[47] C. H. Bennett, P. Hayden, D. W. Leung, P. W. Shor, and A. Winter, IEEE Trans. Inf. Theory **51**, 56 (2005).
[48] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, Phys. Rev. Lett. **78**, 2275 (1997).
[49] E. M. Rains, Phys. Rev. A **60**, 179 (1999).
[50] E. M. Rains, IEEE Trans. Inf. Theory **47**, 2921 (2001).
[51] X. Wang and R. Duan, Phys. Rev. A **95**, 062322 (2017).
[52] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **81**, 5932 (1998).
[53] W. Dür, H.-J. Briegel, J. I. Cirac, and P. Zoller, Phys. Rev. A **59**, 169 (1999).