

Adaptive channel reshaping for improved entanglement distillation

Dina Abdelhadi,^{1,*} Tomas Jochym-O'Connor^{2,3}, Vikesh Siddhu,^{2,4} and John Smolin²

¹*School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland*

²*IBM Quantum, IBM T.J. Watson Research Center, Yorktown Heights, New York, USA*

³*Almaden Research Center, San Jose, California, USA*

⁴*IBM Quantum, IBM Research India, New Delhi, India*



(Received 3 February 2025; accepted 4 November 2025; published 12 January 2026)

Quantum communication and computation heavily rely on entanglement distillation protocols. There is a plethora of distillation protocols for Pauli channels and also for some non-Pauli channels. However, an effort to relate the effectiveness of these protocols for different types of noise has been missing. For most quantum channels, the gap between the existing lower and upper bounds on distillation rates is substantial, and improvements of achievable rates have been stagnant for decades. In this work, we improve the best known distillation lower bounds, for both the amplitude damping and depolarizing channels. We build on a key observation that distillation protocols reshape several uses of a very noisy channel into a better effective channel. We apply this channel processing in an adaptive and recurrent manner. For the amplitude damping channel, our suggested protocol reshapes the channel into an erasure channel, achieving rates exceeding the best known lower bound given by the channel's reverse coherent information. For the depolarizing channel, we introduce the Greedy recurrence protocol with proven performance guarantees and construct a combined protocol improving upon previously known distillation rates. Improved bounds on attainable distillation rates give insights for both practical implementations and theoretical understanding of quantum information processing.

DOI: [10.1103/8kdf-37gp](https://doi.org/10.1103/8kdf-37gp)

I. INTRODUCTION

Useful quantum information processing systems need to share and distribute high-fidelity entangled states across noisy communication links. This link can be a physical interconnect that joins computing or communication modules or a sequence of noisy gates that entangle distant qubits on a single module. A noisy point-to-point link is mathematically modeled by a completely positive trace-preserving map (also called a quantum channel). Entanglement shared across a quantum channel becomes noisy and entanglement distillation is the key protocol to remove this noise [1].

The ultimate benchmark for carrying out entanglement distillation across a quantum channel is its quantum capacity: the number of Bell pairs per channel use that can be shared between channel sender and receiver with asymptotically vanishing error [1–4]. It is customary to assume availability of free noiseless classical messaging for entanglement distillation since the cost of using classical channels is much less than that of a quantum channel.

A channel's quantum capacity depends on whether one can send classical messages one way, from channel sender to receiver, or two way, to and from sender and receiver. In

the former case, we obtain the one-way quantum capacity Q , and in the latter we obtain the two-way quantum capacity Q_2 . These capacities can be very different from one another [1], for instance a qubit erasure channel with erasure probability ϵ has $Q_2 - Q = \min(\epsilon, 1 - \epsilon)$ [5].

A few main goals of quantum Shannon theory are to gain insight into channel capacities, find mathematical expressions for computing them, and protocols to achieve them. One such insight is that the one-way quantum capacity benchmarks quantum error correction while the two-way capacity benchmarks quantum communication assisted by classical communication (see Sec. V in Ref. [1]). While there are well-known expressions for the one-way quantum capacity in terms of a channel's coherent information [6–9], no comparable expressions are known for the two-way quantum capacity. As a result, the field of quantum Shannon theory for quantum communication remains open.

The absence of a capacity expression makes it challenging to find lower bounds on the two-way quantum capacity of general channels. Prior work on such bounds focuses instead on specific states and channels of interest to gain insight.

The earliest asymptotic protocols for entanglement distillation are breeding and hashing [1]. They apply to a mixture of Bell states and proceed by doing random parity checks. The authors of Ref. [10] elevate the breeding protocol by proposing to do an asymptotic version of a single parity check prior to carrying out breeding. These protocols give lower bounds on Q_2 for qubit Pauli channels. For a general channel, the Devetak-Winter [9] protocol achieves a rate for one-way entanglement distillation equaling the channel's coherent information. Finding a truly two-way protocol to

*Contact author: dinaabdelhadiquantum@gmail.com

exceed this rate can be hard. However, simply reversing the role of channel sender and receiver in the Devetak-Winter [9] scheme gives the best known bound for the qubit amplitude damping (AD) channel called the channel's reverse coherent information [9,11–13].

Directly using the breeding or the Ref. [10] scheme does not necessarily result in the best known rates for two-way entanglement distillation. When distilling a mixture of Bell states, a finite-sized protocol prior to an asymptotic protocol can lead to nontrivial improvements in the rate [10,14]. A finite-sized protocol can be selected for matching a ZZ or XX parity check on two noisy two-qubit states (sometimes called the recurrence protocol); however, it can be useful to do several parity checks at once to localize errors across several two-qubit states [15]. Instead of localizing errors on states, Hostens *et al.* [16] come up with a sequence of parity checks aimed at reducing the entropy of the Pauli error distribution.

Progress on finding optimal rates for distillation is crucial. Unfortunately, progress has been slow, for instance, rates for the amplitude damping channel have not been improved in a decade and progress on Pauli channels has been few and far between. Combining finite and asymptotic protocols for distillation can be crucial to making progress; however, the interplay between finite-sized protocols and asymptotic protocols remains poorly understood. For instance, the earliest protocols combining finite-sized recurrence with hashing do not adapt the finite-sized protocol, the Vollbrecht and Verstraete [10], as well as the Hostens-Dahene-De Moor [16] protocols, both require some form of recurrence-based preprocessing to improve on hashing. Yet, the best way to select a finite-sized protocol is not necessarily clear. Given that there are a variety of finite-sized protocols, each developed from a different perspective, there is a need to investigate how these approaches relate to one another and feed into asymptotic schemes for distillation.

Our main contribution is to achieve state-of-the-art rates for (two-way classically assisted) entanglement distillation across the qubit Pauli and amplitude damping channels in the high-noise regime. Key to obtaining these rates is a conceptual insight: Any finite-sized entanglement distillation protocol is a way to reshape one noisy channel into a better channel. That is, upon completion of a finite-sized protocol, the effective error channel on the output state is transformed; specifically, it has been reshaped. This reshaping idea captures the use of error-correcting codes and two-way classical communication to distill entanglement but more importantly it is amenable to concatenation with itself and gives intuition to help combine different asymptotic and finite-sized distillation protocols with each other.

We reshape the AD channel using constant weight encodings, that is, a superposition over computational basis states with a constant number of 1s, to an improved channel and obtain rates that exceed the channel's reverse coherent information. Our simplest dual-rail and triple-rail encodings reshape the damping channel into an erasure channel. Using two-way capacity achieving schemes for the erasure channel, we obtain rates that exceed the AD channel's reverse coherent information. In certain noise regimes, more complex encodings achieving even higher rates use Hamming-weight-2 encodings to reshape the AD channel into a direct sum of

erasure and nonerasure channels. Entanglement from nonerasure channels can be extracted using different protocols (see discussion in Sec. II D).

Similarly to the AD case, the idea of reshaping turns out to be beneficial for thinking about distillation over Pauli channels. We focus on the high-noise regime where rounds of preprocessing using recurrence are typically necessary for achieving nonzero rates. We observe that several rounds of successful recurrence across a Pauli channel give an improved Pauli noise channel. For this channel, a higher-rate parity check is more suitable for detecting errors. Thus, we concatenate a number of rounds of recurrence with a round of parity checks of the $[[4, 2, 2]]$ code. Moreover, we note how each recurrence round reshapes two channel uses into a qubit Pauli channel, where the dominant Pauli noise component changes according to the check performed. Thus, we design a heuristic Greedy recurrence protocol to adapt to this change in the noise distribution, and prove that this heuristic is guaranteed to improve the entanglement fidelity at every round. Using these ideas, we achieve state-of-the-art rates for entanglement distillation over the depolarizing channel.

Notation and preliminaries

We denote the entropy of a discrete probability distribution $[p_i]$ by $h([p_i]) := -\sum_i p_i \log_2 p_i$, the binary entropy function as $h_b(x) := -x \log_2 x - (1-x) \log_2 (1-x)$, and the Von Neumann entropy by $S(\rho) := -\text{Tr}[\rho \log_2 \rho]$. A maximally entangled state of dimension d is denoted as $|\Phi_d\rangle := \sum_{i=0}^{d-1} |ii\rangle / \sqrt{d}$. The Pauli Z matrix is defined as $Z := |0\rangle\langle 0| - |1\rangle\langle 1|$, while $Z_i := \mathbb{I}^{\otimes i-1} \otimes Z \otimes \mathbb{I}^{\otimes n-i}$ denotes Pauli Z applied to the i th qubit. Similarly, the Pauli X matrix is defined as $X := |+\rangle\langle +| - |-\rangle\langle -|$, while $X_i := \mathbb{I}^{\otimes i-1} \otimes X \otimes \mathbb{I}^{\otimes n-i}$, where $|\pm\rangle = (|0\rangle \pm |1\rangle) / \sqrt{2}$. We denote the outcomes of a Z measurement to be +1 and -1 corresponding to eigenstates $|0\rangle$ and $|1\rangle$, respectively. This same ± 1 convention is used to denote outcomes of measuring products of Pauli operators Z_i or X_i . Such product measurements are sometimes called parity measurements.

A quantum channel is a completely positive, trace-preserving map; it maps valid quantum states to valid quantum states. A particularly useful representation of quantum channels is the Kraus representation, where the action of a channel, say $\mathcal{E}$, is represented via its Kraus operators $\{K_i\}$ as $\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger$, where $\sum_i K_i^\dagger K_i = \mathbb{I}$ represents the identity operator and ρ represents a density operator (see Ch. 4 in Ref. [17] or Ch. 2 in Ref. [18] for more details).

The coherent information of a bipartite state ρ_{AB} is given by $I_c(\rho_{AB}) := S(\rho_B) - S(\rho_{AB})$, while the coherent information of a quantum channel $\mathcal{E}_{A' \rightarrow B}$ is given by $I_c(\mathcal{E}) := \max_{\sigma_{AA'}} I_c(\mathcal{I}_A \otimes \mathcal{E}_{A' \rightarrow B}(\sigma_{AA'}))$ [3,6,19]. Meanwhile, the reverse coherent information of a bipartite state ρ_{AB} is given by $I_r(\rho_{AB}) := S(\rho_A) - S(\rho_{AB})$, while the reverse coherent information of a quantum channel $\mathcal{E}_{A' \rightarrow B}$ is given by $I_r(\mathcal{E}) := \max_{\sigma_{AA'}} I_r(\mathcal{I}_A \otimes \mathcal{E}_{A' \rightarrow B}(\sigma_{AA'}))$ [9,11–13]. We will be studying the problem of entanglement distillation (sometimes referred to as entanglement sharing, entanglement generation, or entanglement purification). Entanglement distillation may refer to the setting where two parties share n copies of a noisy entangled state and attempt to distill noiseless entanglement out of these

copies, or to the setting where the communicating parties are connected by a noisy quantum channel that they use n times with the goal of sharing noiseless entanglement at the conclusion of the protocol. In this work, we focus on the latter meaning, though the two interpretations somewhat overlap. In particular, following Ref. [8], in the entanglement sharing and distillation protocols we consider, Alice and Bob employ n uses of a noisy quantum channel with the goal of ending up with a shared state ρ such that $\langle \Phi_{d_n} | \rho | \Phi_{d_n} \rangle \geq 1 - \epsilon_n$, and $\epsilon_n \rightarrow 0$, as $n \rightarrow \infty$. For such protocols, the distillation rate is defined as $R = \lim_{n \rightarrow \infty} \log_2(d_n)/n$. The term two-way assisted entanglement distillation refers to protocols designed to accomplish the above task, where classical communication between Alice and Bob in both directions is treated as a free resource throughout the rounds of the protocol.

II. AMPLITUDE DAMPING CHANNEL

The AD channel is one of the most well-studied examples of non-Pauli channels. It is a well-motivated and realistic noise model, representing the decay of a quantum system from an excited state $|1\rangle$ to a low-energy (ground) state $|0\rangle$. We denote the amplitude damping channel with damping parameter γ by $\mathcal{A}_\gamma$. The Kraus operators of $\mathcal{A}_\gamma$ are given by

$$A_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix}, \quad A_1 = \begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix}. \quad (1)$$

In the unassisted quantum communication setting, the amplitude damping channel is one of a few rare examples where the capacity is exactly known. The amplitude damping channel is degradable when $0 \leq \gamma < 1/2$ and antidegradable otherwise [20,21]. Thus, its unassisted capacity is equal to its coherent information for $0 \leq \gamma < 1/2$ and 0 otherwise.

In the two-way assisted quantum communication setting over the amplitude damping channel, the optimal rates are not known (see Ref. [22] for bounds on the capacity). The reverse coherent information has long been considered to be the best known lower bound on the two-way assisted quantum capacity of the amplitude damping channel [13]. Using the concavity of the reverse coherent information in terms of the input state, the authors of Ref. [13] obtain the following expression for the reverse coherent information of the amplitude damping channel:

$$I_r(\mathcal{A}_\gamma) = \max_{0 \leq x \leq 1} h_b(x) - h_b(\gamma x). \quad (2)$$

A. Dual-rail encoding

The dual-rail encoding has been suggested as a practical way of converting amplitude damping errors into erasure errors [23,24], which are in turn easier to handle and correct. In this encoding, an input logical qubit state is encoded into two physical qubits, according to the following mapping:

$$|0\rangle \rightarrow |01\rangle, \quad |1\rangle \rightarrow |10\rangle.$$

Considering all the possible amplitude damping errors applied to the two physical qubits, we have

$$\begin{aligned} A_0 \otimes A_0 |01\rangle &= \sqrt{1-\gamma} |01\rangle, & A_0 \otimes A_0 |10\rangle &= \sqrt{1-\gamma} |10\rangle, \\ A_0 \otimes A_1 |01\rangle &= \sqrt{\gamma} |00\rangle, & A_1 \otimes A_0 |10\rangle &= \sqrt{\gamma} |00\rangle. \end{aligned}$$

Note that $A_1 \otimes A_0 |01\rangle = A_1 \otimes A_1 |01\rangle = A_0 \otimes A_1 |10\rangle = A_1 \otimes A_1 |10\rangle = \mathbf{0}$; i.e., the damping operator A_1 annihilates $|0\rangle$ to the zero vector $\mathbf{0}$ as $|0\rangle$ is already in the lowest level. Thus, the input encoded state is mapped by the amplitude damping channel either back to the codespace or to an orthogonal subspace spanned by $|00\rangle$, which plays the role of the erasure state, as it is orthogonal to the codespace. The output state after encoding and transmission through the amplitude damping channel is

$$\mathcal{A}_\gamma^{\otimes 2}(\rho_{\text{encoded}}) = (1-\gamma)\rho_{\text{encoded}} + \gamma|00\rangle\langle 00|.$$

By using code concatenation along with the dual-rail encoding, authors of Ref. [23] construct multi-error-correcting codes.

In the two-way assisted quantum communication setting, we now show that it suffices to use the dual-rail encoding for erasure conversion or equivalently, for amplitude damping error detection rather than correction to achieve rates higher than the reverse coherent information for some parameters of γ .

Protocol. Alice encodes half of an entangled Bell pair into the dual-rail encoding, and sends the encoded qubits to Bob across the amplitude damping channel, employing two channel uses per entangled pair. The overall encoded state is

$$|\Phi_2\rangle_{\text{encoded}} = \frac{|0\rangle_A |01\rangle_{A'} + |1\rangle_A |10\rangle_{A'}}{\sqrt{2}},$$

where Alice keeps the system A and transmits the system A' over the channel to Bob. Simply by measuring the parity $Z_{B_1} \otimes Z_{B_2}$ of the two qubits he receives, Bob can determine if the received state is intact or has been subjected to a damping error, mapping it to the erasure state $|00\rangle$. The rest of the protocol proceeds similarly to the simple protocol for two-way assisted quantum communication over the quantum erasure channel. If Bob detects the erasure state $|00\rangle$, he uses classical communication to tell Alice that the shared encoded entangled pair should be discarded. Otherwise, the shared entangled pair is kept and is error free. The shared encoded pair is kept with probability $1-\gamma$. As we can extract at most one entangled pair per two channel uses, we normalize by the number of channel uses, which represents a resource overhead. This normalization per channel use is applied to all the computations of protocol yields in this work. Thus, the dual-rail encoding has a rate of $1/2$, and the overall yield of this protocol is

$$Y_{\text{dual-rail}}(\gamma) = \frac{1-\gamma}{2}.$$

Take $\gamma = 2/3$, then, $Y_{\text{dual-rail}}(2/3) = 1/6 > 0.16666$, while $I_r(\mathcal{A}_{\gamma=2/3}) < 0.16148$ (see Fig. 1).

It is shown in Ref. [25] that using the dual-rail encoding, the two-way assisted entanglement sharing rate can exceed that achievable by the reverse coherent information for the damping-dephasing channel. This improvement can be explained by the fact that this encoding can be used to isolate away the damping noise in the damping-dephasing channel.

B. Extension to triple-rail encoding

A generalization of the dual-rail encoding for error correction suggested in Ref. [23] can be adapted for improving asymptotic rates for two-way assisted entanglement distillation. Consider an input logical qutrit state, with basis

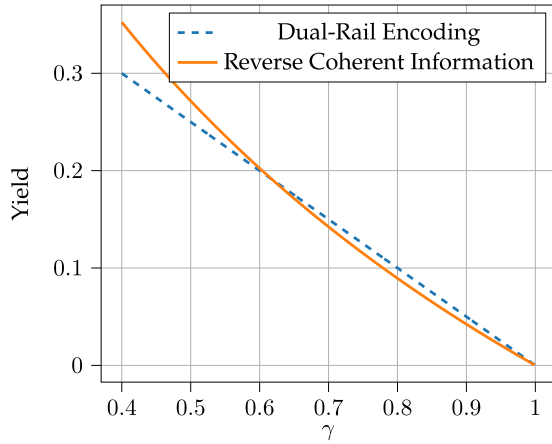


FIG. 1. Comparison between the yield of the dual-rail encoding scheme and the reverse coherent information showing a range where the dual-rail encoding outperforms the reverse coherent information.

states $\{|0\rangle_L, |1\rangle_L, |2\rangle_L\}$, encoded into three physical qubits according to the mapping

$$|0\rangle_L \rightarrow |001\rangle, \quad |1\rangle_L \rightarrow |010\rangle, \quad |2\rangle_L \rightarrow |100\rangle.$$

Using a similar protocol to the dual-rail case, Alice and Bob can distill $(\log_2 3)/3$ per channel use with probability $1 - \gamma$, giving a strictly larger yield than the dual-rail case,

$$Y_{\text{triple-rail}}(\gamma) = \frac{(1 - \gamma) \log_2 3}{3} > Y_{\text{dual-rail}}(\gamma). \quad (3)$$

Protocol. Alice encodes one half of a three-dimensional entangled state $|\Phi_3\rangle$, obtaining the following encoded state:

$$|\Phi_3\rangle_{\text{encoded}} = \frac{|0\rangle_A |001\rangle_{A'} + |1\rangle_A |010\rangle_{A'} + |2\rangle_A |100\rangle_{A'}}{\sqrt{3}}.$$

Alice sends the A' qubits to Bob over three uses of the amplitude damping channel, with the resulting overall state being

$$\begin{aligned} \mathcal{I}_A \otimes \mathcal{A}_\gamma^{\otimes 3}(\Phi_{3\text{encoded}}) &= (1 - \gamma) |\Phi_3\rangle \langle \Phi_3|_{\text{encoded}, AB} \\ &\quad + \gamma \frac{\mathbb{I}_A}{3} \otimes |000\rangle \langle 000|_B. \end{aligned}$$

Bob can then detect whether the three received qubits are in the erasure state $|000\rangle$ or in the code space by measuring the parity of the received qubits. They can do so by measuring $Z_{B_1} Z_{B_2} Z_{B_3}$ and only accepting if the outcome of the measurement is -1 . Overall, with probability $(1 - \gamma)$ Alice and Bob can distill $\log_2 3$ ebits over three channel uses.

C. Further Hamming-weight-1 encodings

One may wonder if extending this encoding to n channel uses may yield even better rates. Consider the following encoding of half of an n -dimensional entangled state:

$$\begin{aligned} |\Phi_n\rangle &= \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} |i\rangle_A |i\rangle_{A'} \rightarrow |\Phi_n\rangle_{\text{encoded}} \\ &= \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} |i\rangle_A \otimes X_i |0\rangle_{A'}^{\otimes n}. \end{aligned}$$

Let $A_x := \bigotimes_{i=0}^{n-1} A_{x_i}$, $x \in \{0, 1\}^n$ be the tensor product of the amplitude damping channel's Kraus operators, $\{A_0, A_1\}$ defined in Eq. (1), and

$$A_k := A_0^{\otimes n-k-1} \otimes A_1 \otimes A_0^{\otimes k}. \quad (4)$$

Sending the A' systems over n uses of the amplitude damping channel yields the following state shared between Alice and Bob:

$$\begin{aligned} \mathcal{I}_A \otimes \mathcal{A}_\gamma^{\otimes n}(\Phi_{n,\text{encoded}}) &= \frac{1}{n} \sum_{x \in \{0,1\}^n} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} |i\rangle \langle j|_A \otimes A_x X_i |0\rangle \langle 0|_B^{\otimes n} X_j A_x^\dagger \\ &= \frac{1}{n} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} |i\rangle \langle j|_A \otimes A_0^n X_i |0\rangle \langle 0|_B^{\otimes n} X_j A_0^n \\ &\quad + \frac{1}{n} \sum_{i=0}^{n-1} |i\rangle \langle i|_A \otimes A_i X_i |0\rangle \langle 0|_B^{\otimes n} X_i A_i^\dagger \\ &= (1 - \gamma) |\Phi_n\rangle \langle \Phi_n|_{\text{encoded}} + \gamma \frac{\mathbb{I}_A}{n} \otimes |0\rangle \langle 0|_B^{\otimes n}. \end{aligned}$$

Thus, by measuring the parity $Z^{\otimes n}$, Bob can determine if the received state has been “erased” into the state $|0\rangle^{\otimes n}$ or not. This way, by using the amplitude damping channel n times, Alice and Bob can distill $\log_2 n$ shared ebits with probability $1 - \gamma$, thus giving a scheme with yield

$$Y_{\text{Hamming-1}}(n, \gamma) = \frac{(1 - \gamma) \log_2 n}{n}. \quad (5)$$

This expression shows that the family of Hamming-weight-1 codes is *not* a group of constant rate codes, with a sublinear logical dimension relative to the blocklength; i.e., larger codes incur a higher encoding cost. Thus, the rate is maximized for some finite value of n .

Lemma 1. Let $n > 0$ be an integer, then $(\log_2 n)/n$ is maximized at $n = 3$.

Proof. At $n = 1$, $\log_2 1 = 0$, at $n = 2$, $(\log_2 2)/2 = 1/2$, and at $n = 3$, $(\log_2 3)/3 > 0.528$.

For $n \geq 3$,

$$\begin{aligned} \frac{\log_2(n+1)}{n+1} - \frac{\log_2 n}{n} &= \frac{\log_2 \left[\left(1 + \frac{1}{n}\right)^n \right] - \log_2 n}{n(n+1)} \\ &\leq \frac{\log_2 \frac{e}{n}}{n(n+1)} < 0. \end{aligned}$$

Thus, using Lemma 1, we see that the Hamming-weight-1 encodings already achieve their maximum yield at $n = 3$ with the triple-rail encoding.

D. Hamming-weight-2 encodings

We now consider the case of encodings of Hamming weight 2 over n qubits. For simplicity, we define the following notation:

$$|a, b\rangle := X_a X_b |0\rangle^{\otimes n} \quad \text{and} \quad |a\rangle := X_a |0\rangle^{\otimes n}, \quad (6)$$

where integers a and b lie between 0 and $n - 1$, inclusive. The number of binary strings of length n and Hamming weight 2 is $m = \binom{n}{2} = n(n - 1)/2$.

Protocol. Alice encodes an m -dimensional entangled state as follows:

$$\begin{aligned} |\Phi_m\rangle &= \frac{1}{\sqrt{m}} \sum_{i=0}^{m-1} |i\rangle_A |i\rangle_{A'} \rightarrow |\Phi_m\rangle_{\text{encoded}} \\ &= \frac{1}{\sqrt{m}} \sum_{a=0}^{n-2} \sum_{b>a}^{n-1} |a, b\rangle_A \otimes |a, b\rangle_{A'}. \end{aligned}$$

Alice then sends the A' systems to Bob over n uses of the amplitude damping channel. Note that of the Kraus operators of n applications of the amplitude damping channel to the encoded state, only those with at most two occurrences of the A_1 operator survive, due to the fact that $A_1|0\rangle = \mathbf{0}$. Using the notation from Eq. (4), the resulting shared state can be expressed as

$$\begin{aligned} \mathcal{I}_A \otimes \mathcal{A}_\gamma^{\otimes n}(\Phi_{m,\text{encoded}}) &= (1-\gamma)^2 \Phi_{m,\text{encoded}} + \gamma^2 \frac{1}{m} \sum_{\substack{a, \\ b>a}} |a, b\rangle \langle a, b| \otimes |0\rangle \langle 0|^{\otimes n} \\ &\quad + \underbrace{\frac{1}{m} \sum_{\substack{a, \\ b>a}} \sum_{\substack{a', \\ b'>a'}} \sum_{k=0}^{n-1} |a, b\rangle \langle a', b'| \otimes \mathbb{A}_k |a, b\rangle \langle a', b'| \mathbb{A}_k^\dagger}_{\rho_1}. \end{aligned}$$

Focusing on the state ρ_1 that results from a single A_1 damping event, and noting that $\mathbb{A}_k |a, b\rangle = \mathbf{0}$ unless $k = a$ or $k = b$, we get

$$\begin{aligned} \rho_1 &= \frac{\gamma(1-\gamma)}{m} \left[\sum_{k, b>k, b'>k} |k, b\rangle \langle k, b'| \otimes |b\rangle \langle b'| + \sum_{k, a<k, a'<k} |a, k\rangle \langle a', k| \otimes |a\rangle \langle a'| \right. \\ &\quad \left. + \sum_{k, b>k, a'<k} |k, b\rangle \langle a', k| \otimes |b\rangle \langle a'| + \sum_{k, a<k, b'>k} |a, k\rangle \langle k, b'| \otimes |a\rangle \langle b'| \right] \\ &= \frac{\gamma(1-\gamma)}{m} \left[\sum_{k, b \neq k, b'>k} |k, b\rangle \langle k, b'| \otimes |b\rangle \langle b'| + \sum_{k, a \neq k, a'<k} |a, k\rangle \langle a', k| \otimes |a\rangle \langle a'| \right] \\ &= \frac{\gamma(1-\gamma)}{m} \left[\sum_{k, b \neq k, b' \neq k} |k, b\rangle \langle k, b'| \otimes |b\rangle \langle b'| \right] \\ &= 2\gamma(1-\gamma) \frac{1}{n} \sum_k |v_k\rangle \langle v_k|, \end{aligned}$$

where $|v_k\rangle = (n-1)^{-1/2} \sum_{b=0: b \neq k}^{n-1} |b, k\rangle_A |b\rangle_B$. The overall state is then

$$\mathcal{I}_A \otimes \mathcal{A}_\gamma^{\otimes n}(\Phi_{m,\text{encoded}}) = (1-\gamma)^2 \Phi_{m,\text{encoded}} + \gamma^2 \frac{1}{m} \sum_{\substack{a, \\ b>a}} |a, b\rangle \langle a, b| \otimes |0\rangle \langle 0|^{\otimes n} + 2\gamma(1-\gamma) \frac{1}{n} \sum_k |v_k\rangle \langle v_k|. \quad (7)$$

Bob can distinguish the case of an even number (0 or 2) of A_1 errors from the case of a single A_1 error by measuring the check $Z_B^{\otimes n}$ on their qubits. If the measurement outcome is -1 , then a single error has occurred. The outcome -1 is obtained with probability $2\gamma(1-\gamma)$. If the outcome is 1, Bob can further distinguish the case of total erasure (two A_1 errors) from the case of no A_1 errors by applying the projective measurement with the projectors $\{|0\rangle\langle 0|^n, \mathbb{I} - |0\rangle\langle 0|^{\otimes n}\}$. In the case of no A_1 errors, $\log_2 m$ shared ebits can be distilled, as the original state is unaffected. This occurs with probability $(1-\gamma)^2$. In the case of a single A_1 error, the state shared between Alice and Bob is $\sigma_{AB} = \sum_{k=0}^{n-1} |v_k\rangle \langle v_k|/n$, with $\langle v_k | v_{k'} \rangle = \delta_{k, k'}$. Thus, it is a state with n eigenvalues, each being $1/n$. The state $\sigma_A = \text{Tr}_{A'}[\Phi_{m,\text{encoded}}]$ is up to an isometry $= \mathbb{I}_m/m$, as Alice's systems did not go through the channel. When Bob signals to Alice that the state they received suffered from a single damping error, they can then proceed to apply the protocol for entanglement distillation that achieves the reverse

coherent information [13]. The reverse coherent information of this state is

$$\begin{aligned} I_r(\sigma_{AB}) &= S(\sigma_A) - S(\sigma_{AB}) = \log_2(m) - \log_2(n) \\ &= \log_2 \frac{n-1}{2}. \end{aligned}$$

The overall yield of the scheme is

$$\begin{aligned} Y_{\text{Hamming-2}}(n, \gamma) &= \frac{1}{n} \left[(1-\gamma)^2 \log_2 \frac{n(n-1)}{2} \right. \\ &\quad \left. + 2\gamma(1-\gamma) \log_2 \frac{n-1}{2} \right]. \end{aligned}$$

In the case of $n = 4$ and $m = 6$, we can do slightly better than to pass the state σ_{AB} onto the reverse coherent strategy, which achieves $\log_2[(n-1)/2]$. In this case, we may rewrite

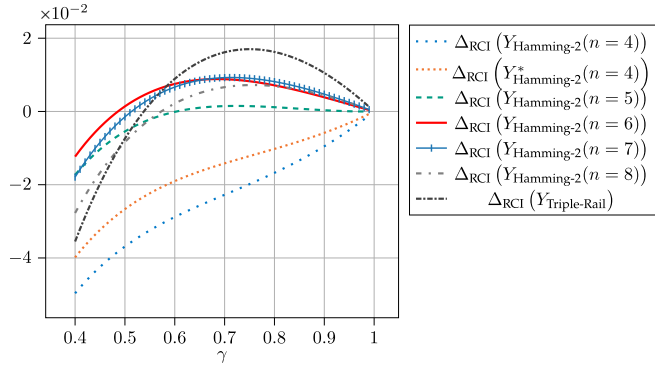


FIG. 2. Comparison of the improvements of yields of various Hamming-weight-2 encodings, and the triple-rail encoding, over the reverse coherent information rate. For any yield expression Y , let $\Delta_{\text{RCI}}(Y) = Y - I_r(\mathcal{A}_\gamma)$.

the vectors $|v_k\rangle$ in the computational basis as follows:

$$\begin{aligned}
 |v_0\rangle &= \frac{1}{\sqrt{3}}(|1100\rangle|0100\rangle + |1010\rangle|0010\rangle + |1001\rangle|0001\rangle), \\
 |v_1\rangle &= \frac{1}{\sqrt{3}}(|1100\rangle|1000\rangle + |0110\rangle|0010\rangle + |0101\rangle|0001\rangle), \\
 |v_2\rangle &= \frac{1}{\sqrt{3}}(|1010\rangle|1000\rangle + |0110\rangle|0100\rangle + |0011\rangle|0001\rangle), \\
 |v_3\rangle &= \frac{1}{\sqrt{3}}(|1001\rangle|1000\rangle + |0101\rangle|0100\rangle + |0011\rangle|0010\rangle).
 \end{aligned} \tag{8}$$

Once Alice and Bob know that a single damping error event occurred, then Bob measures $Z_0 \otimes Z_1$, and communicates the outcome to Alice. If the error commutes with $Z_0 \otimes Z_1$, then Alice measures each of their first two qubits in the computational basis, obtaining a two-bit binary outcome b_0, b_1 . If $b_0 \oplus b_1 = 0$, then Alice and Bob discard the state. Otherwise, they have gained a Bell pair. If the error anticommutes with $Z_0 \otimes Z_1$, then Alice measures each of their second and third qubits in the computational basis, obtaining a two-bit outcome b_0, b_1 . If $b_0 \oplus b_1 = 0$, then Alice and Bob discard the state. Otherwise, they have gained a Bell pair. Thus, overall with probability $2/3$, Alice and Bob can distill a Bell pair, in the case of a single damping event, which gives the yield

$$Y_{\text{Hamming-2}}^*(4, \gamma) = \frac{1}{4}[(1 - \gamma)^2 \log_2 6 + 2\gamma(1 - \gamma)^{\frac{2}{3}}].$$

Although $Y_{\text{Hamming-2}}^*(4, \gamma) > Y_{\text{Hamming-2}}(4, \gamma)$, the rate of the original reverse coherent information protocol still exceeds both. It is an open question to devise similar improved protocols for $n > 4$. From the plots of Figs. 2–4, we see that an advantage over the reverse coherent information yield [Eq. (2)] can be obtained at $n = 6$.

III. PAULI CHANNELS

In Sec. III A, we define notation that we will use to discuss distillation over Pauli channels and Sec. III B discusses prior work of Refs. [10,14,15]. Readers familiar with these standard

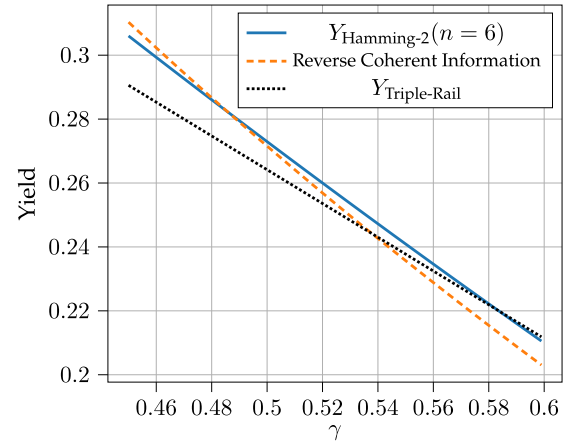


FIG. 3. Comparison of the yields of the Hamming-weight-2 encoding at $n = 6$, Reverse Coherent Information, and Triple-Rail Encoding.

discussions may skip ahead to Sec. III C, where we propose protocols.

A. Notation for entanglement distillation over Pauli channels

The four Bell states are denoted as $|B_{x,z}\rangle = I \otimes X^x Z^z(|00\rangle + |11\rangle)/\sqrt{2}$, where $x, z \in \{0, 1\}$. Note that $|B_{00}\rangle = |\Phi_2\rangle$. Graphically, Bell states are denoted as in Fig. 5, with the upper wire indicating the half of the Bell pair kept by Alice, and the lower one indicating the half sent over the noisy channel to Bob.

Let the Pauli matrices be denoted by $\mathcal{P} = \{I, X, Y, Z\}$. A capital letter subscript attached to a Pauli operator is used to indicate the subsystem of a given Pauli operator.

As we will be referring to stabilizer codes, we give a brief recap of stabilizer codes here. Following Refs. [26,27], a stabilizer code is defined via a stabilizer group S , an Abelian subgroup of $\mathcal{P}^n$, the group of n -qubit Pauli matrices. The code words of such a code $\{|\psi_i\rangle\}$ are simultaneous eigenstates of the elements of S , with a $+1$ eigenvalue. When a code word

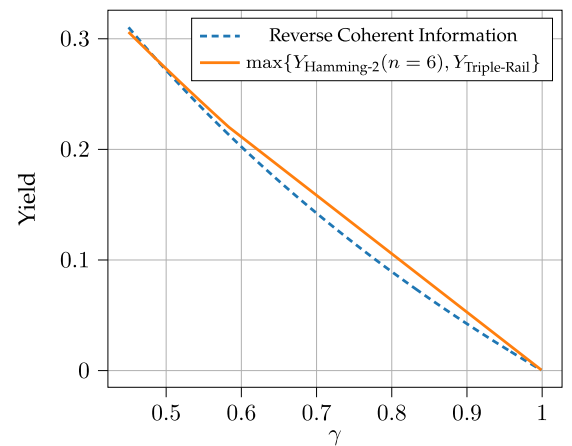
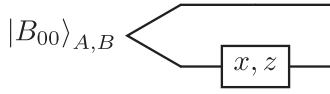


FIG. 4. Comparison of the maximum of the yields of the Hamming-weight-2 encoding at $n = 6$ and the Triple-Rail Encoding vs the Reverse Coherent Information.

FIG. 5. $|B_{x,z}\rangle$.

is sent through an n -qubit Pauli channel, it is affected by Pauli error $E \in \mathcal{P}^n$. Any such Pauli error either commutes or anticommutes with an element $S_j \in S$. Determining whether E commutes or anticommutes with every generator of S gives a binary string called an error syndrome. The syndrome information makes error detection, decoding, and correction possible. The bits of the syndrome can be determined by measuring the stabilizer observables via projective measurements. For each such observable S_j , $\{\Pi_+ = (\mathbb{I} + S_j)/2, \Pi_- = (\mathbb{I} - S_j)/2\}$ define elements of a projective measurement that allows determining the commutation relation of E and S_j .

A Pauli string of length n is a tensor product of Pauli matrices acting on n qubits. Up to a global phase, such a Pauli string can be represented by a binary string of length $2n$. In a similar manner, an n -pair Bell state is denoted by $|B_s\rangle = |B_{x_1,z_1}\rangle \otimes |B_{x_2,z_2}\rangle \cdots \otimes |B_{x_n,z_n}\rangle$, where $s = (x_1, z_1), \dots, (x_n, z_n)$ is a binary string of length $2n$. This mapping between Paulis and binary strings is called the symplectic representation. A more detailed discussion of the symplectic representation is available in Appendix A 3 and a more detailed exposition can be found in Ref. [28]. This representation establishes a direct connection between linear parity checks applied to binary strings and stabilizer measurements. Due to this connection between stabilizer measurements and binary checks, we can denote the results of stabilizer check measurements either by binary bits $\{0, 1\}$, with 0 signaling that the measured stabilizer commutes with the error and 1 indicating anticommutation, or equivalently by $\{+1, -1\}$ with $+1$ indicating commutation and -1 signaling anticommutation as mentioned in Sec. I A.

A Pauli noise channel with probabilities $[p_I, p_X, p_Y, p_Z]$ acts on an input state as follows:

$$\rho \rightarrow p_I \rho + p_X X \rho X + p_Y Y \rho Y + p_Z Z \rho Z. \quad (9)$$

An important example of a Pauli noise channel is the *depolarizing channel*, where $p_X = p_Y = p_Z = p/3$ and $p_I = 1 - p$. An n -qubit quantum state is said to be *Bell diagonal* if it can be expressed as $\rho = \sum_{s \in \{0,1\}^{2n}} p_s |B_s\rangle \langle B_s|$ for some probability

distribution $\{p_s\}$. When a Pauli channel acts on each qubit of a Bell diagonal state, the resulting state is still Bell diagonal.

Useful operations on Bell pairs. In Figs. 6(a) and 6(b), we review some useful operations on Bell pairs.

$$\begin{aligned} |B_{x_1,z_1}\rangle \otimes |B_{x_2,z_2}\rangle &\xrightarrow{\text{Bilateral XOR [Fig.6(a)]}} |B_{x_1,z_1 \oplus z_2}\rangle \otimes |B_{x_1 \oplus x_2,z_2}\rangle, \\ |B_{x,z}\rangle &\xrightarrow{H^{\otimes 2} [\text{Fig.6(b)]}} |B_{z,x}\rangle. \end{aligned}$$

Entanglement fidelity. The entanglement fidelity of a two-qubit quantum state ρ was defined in Ref. [29] as $F_e(\rho) \equiv \langle B_{00} | \rho | B_{00} \rangle$. If the state is Bell diagonal, i.e.,

$$\begin{aligned} \rho = & p_I |B_{00}\rangle \langle B_{00}| + p_X |B_{10}\rangle \langle B_{10}| + p_Y |B_{11}\rangle \langle B_{11}| \\ & + p_Z |B_{01}\rangle \langle B_{01}|, \end{aligned} \quad (10)$$

then the entanglement fidelity is p_I .

B. Literature review

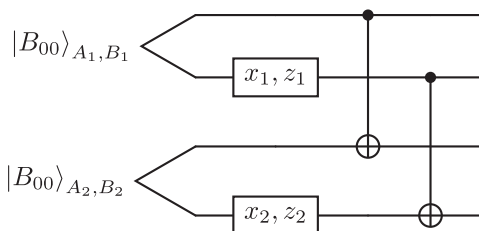
Pauli noise is one of the most widely studied noise models in quantum information, with a lot of attention being devoted to studying the depolarizing noise model. Entanglement sharing over Pauli channels has been studied in several works, with some of the main ideas established in the works of Refs. [10,14,15]. In all of these works, the protocols can be broadly described as follows:

(1) Alice sends halves of Bell pairs to Bob across a noisy Pauli channel, resulting in a Bell-diagonal state shared between Alice and Bob.

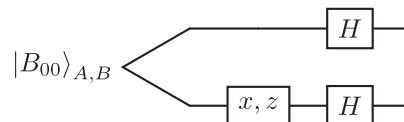
(2) Alice and Bob use the resources available to them (one-way or two-way local operations and classical communication) to measure whether the Pauli error that occurred commutes or anticommutes with a certain Pauli string or stabilizer.

(3) Based on the outcome, Alice and Bob may decide to measure another check, or to apply a Pauli correction to their state, or to discard some qubits.

(4) At the end of the protocol's iterations, Alice and Bob have used the channel n times to end up with a shared state ρ , such that $\langle B_{00}^{\otimes k} | \rho | B_{00}^{\otimes k} \rangle = 1 - \epsilon_n$, with $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$, achieving a distillation rate of $\lim_{n \rightarrow \infty} k(n)/n$, where the notation $k(n)$ indicates that the number of distilled entangled pairs k is a function of the number of channel uses n .



(a) Bilateral XOR applied after Pauli noise, x, z denotes the application of the Pauli error $X^x Z^z$.



(b) Applying a Hadamard gate to each qubit of a noisy Bell pair.

FIG. 6. Useful operations on Bell pairs.

Doing parity checks. We consider the setting where Alice inputs n halves of Bell pairs into the Pauli noise channel, and the resulting shared state is a Bell-diagonal state of the form

$$\begin{aligned}\rho_n &= \sum_{s \in \{0,1\}^{2n}} p_s |B_s\rangle\langle B_s| \\ &= \sum_{P \in \mathcal{P}^n} \Pr[P_B](\mathbb{I}_A \otimes P_B) |B_{00}\rangle\langle B_{00}|^{\otimes n} (\mathbb{I}_A \otimes P_B). \quad (11)\end{aligned}$$

Consider two Pauli strings $P_B, Q \in \mathcal{P}^n$, where P_B is a Pauli acting on Bob's part of the state and Q is a stabilizer check. Alice and Bob share a state $|\psi\rangle_{AB} = \mathbb{I}_A \otimes P_B |B_{00}\rangle_{AB}^{\otimes n}$ and wish to determine whether P_B and Q commute or anticommute. This information can be determined in two possible ways. The first is that Alice and Bob carry out local measurements directly on their respective parts of the shared state, consuming one of the shared noisy Bell pairs, and comparing their outcomes via classical communication to deduce the parity. The second is that they carry out local measurements assisted by a shared ancillary perfect Bell pair, consuming this ancillary pair in the process, then using classical communication to compute the resulting parity. According to the terminology used in Ref. [16], the former would be called a Bilateral Pauli Measurement (BPM), while the latter is called an appended ebit measurement (AEM). For a self-contained discussion, we provide a summary regarding AEMs and BPMs in Appendix B 1.

Finite and Infinite-Size Checks. We are interested in optimizing the asymptotic rates achievable for entanglement distillation. In this case, we assume Alice and Bob share n noisy Bell pairs in the state $\sum_{s \in \{0,1\}^{2n}} p_s |B_s\rangle\langle B_s|$, where $n \mapsto \infty$. The checks they apply to their shared state can be either finite-size checks, measuring a small block size, or infinite-size checks; that is, they may scale with system size.

For a finite-size parity check, Alice and Bob agree on a check with symplectic representation r and a finite integer k , which defines the number of Bell pairs involved in a single instance of the check. In terms of the binary symplectic notation, r is a binary string of length $2k$. Denote the state of k of the shared noisy Bell pairs by $\sum_{t \in \{0,1\}^{2k}} p_t |B_t\rangle\langle B_t|$. Alice and Bob wish to determine the value of the check given by $f_r(t) = t \cdot r \bmod 2$, for $t, r \in \{0,1\}^{2k}$. To do so, Alice and Bob apply local Cliffords and local Pauli measurements on their respective halves of the noisy state. The measurement results they individually obtain are random, but the correlations of the measurements contain the information about the value of the check. By using classical communication to exchange the measurement outcomes, they can determine $f_r(t)$. Since we are concerned with the asymptotic setting, Alice and Bob group their infinite string of n noisy shared Bell pairs into batches of finite size k , and determine the value of the parity check for each batch.

An infinite-size check, on the other hand, involves all of the shared pairs at once, where it determines the value of some $f_x(s) = x \cdot s \bmod 2$, for $s, x \in \{0,1\}^{2n}$, as $n \rightarrow \infty$.

With these concepts in mind, we now devote the next few subsections to reviewing some of the most important protocols

proposed in the literature for entanglement distillation over Pauli channels. These protocols will be particularly relevant as building blocks of the combined protocol, which we will construct in following sections. For the sake of a self-contained discussion, we discuss these protocols in more detail in Appendix B.

1. Breeding and hashing

Given a shared noisy state of the form (11), both the breeding and hashing protocols proposed in Ref. [14] rely on carrying out many random parity checks until the binary string s can be determined. Once s is known, Bob can apply a local Pauli correction to obtain noiseless Bell pairs. Under the assumption that the noisy state is the result of n independent and identically distributed (i.i.d.) applications of a Pauli noise channel with probabilities $[p_I, p_X, p_Y, p_Z]$ [Eq. (9)], the authors of Ref. [14] use the asymptotic equipartition property to show that $nh([p_I, p_X, p_Y, p_Z])$ parity checks would suffice to determine s , since each check rules out half of the possible Pauli strings. Thus, hashing and breeding protocols achieve a rate

$$Y_{\text{Hashing}} = 1 - h([p_I, p_X, p_Y, p_Z]). \quad (12)$$

More generally, for a Bell-diagonal state (11) the breeding/hashing style protocols achieve the coherent information of ρ , $n - S(\rho) = n - h(\{p_s\})$ [9].

The difference between breeding and hashing is that the former uses AEMs, while the latter uses BPMs. It is worth noting that such protocols require no backward communication from Bob to Alice.

2. Recurrence

The Z-recurrence protocol was introduced in Ref. [14] as a preprocessing step preceding the hashing protocol, where a BPM finite-size check is applied in a recursive manner. The input of one step of the recurrence protocol is pairs of noisy Bell pairs $I_A \otimes P_{B_1, B_2} |B_{00}\rangle_{A_1 B_1} \otimes |B_{00}\rangle_{A_2 B_2}$. All pairs affected by any two-qubit Pauli error P_{B_1, B_2} that anticommutes with $Z_{B_1} Z_{B_2}$ are rejected. Note that this is exactly the stabilizer of a size 2 Z-type repetition code. The resulting pair, if accepted, has an induced Pauli channel acting on it. Each step of recurrence halves the rate, as one out of every two Bell pairs is consumed by the measurement. One can repeat the recurrence iterations to further improve the induced channel, or switch over to hashing, depending on which yields a better overall rate. Recurrence inherently requires backward communication from Bob to Alice to communicate which Bell pairs are to be accepted, and which are to be discarded. For the depolarizing channel combining recurrence with hashing, the authors of Ref. [14] show that two-way communication rates are provably higher than those possible with one-way communication only. Analysis in Ref. [16] suggests that the power of recurrence is due to entropy reduction, where degenerate Pauli errors (i.e., an error P_{B_1, B_2} , and $Z_{B_1} Z_{B_2} P_{B_1, B_2}$) are mapped together in the induced channel. The best rates shown in Ref. [14] are achieved using the Macchiavello recurrence protocol, discussed in Sec. 3.2.1 of Ref. [14], where rounds of Z recurrence are applied, with each round being followed by a B_x rotation.

3. Interpolation of recurrence and hashing protocol [10]

In Ref. [10], the authors introduce the interpolation of recurrence and hashing protocol, achieving rates exceeding Y_{Hashing} [Eq. (12)] for the depolarizing channel. This advantage of Ref. [10] over hashing is observed at all depolarizing noise rates. This protocol utilizes two-way classical communication between Alice and Bob.

$$\begin{aligned}\rho_{\text{even}} &= \frac{1}{p_{\text{even}}} \sum_{P \in \mathcal{P}_{\text{even}}} \Pr[P_B](\mathbb{I}_A \otimes P_B)|B_{00}\rangle\langle B_{00}|^{\otimes 2}(\mathbb{I}_A \otimes P_B), \\ \rho_{\text{odd}} &= \frac{1}{p_{\text{odd}}} \sum_{P \in \mathcal{P}_{\text{odd}}} \Pr[P_B](\mathbb{I}_A \otimes P_B)|B_{00}\rangle\langle B_{00}|^{\otimes 2}(\mathbb{I}_A \otimes P_B),\end{aligned}\quad (13)$$

with probability $p_{\text{even}} = \sum_{P \in \mathcal{P}_{\text{even}}} \Pr[P_B]$ and $p_{\text{odd}} = \sum_{P \in \mathcal{P}_{\text{odd}}} \Pr[P_B]$, respectively. If the even output is obtained, then the state ρ_{even} is passed onto the hashing protocol. Otherwise, the BPM check ZI is measured, and either the state

$$\begin{aligned}\rho_{\text{odd},0} &= \frac{p_{ZX} + p_{IX}}{p_0} X_B |B_{00}\rangle\langle B_{00}| X_B \\ &\quad + \frac{p_{ZY} + p_{IY}}{p_0} Y_B |B_{00}\rangle\langle B_{00}| Y_B\end{aligned}$$

is obtained with probability p_0/p_{odd} and then further passed onto the hashing protocol, where $p_0 = p_{ZX} + p_{IX} + p_{ZY} + p_{IY}$, or the state

$$\rho_{\text{odd},1} = \frac{p_{XZ} + p_{YZ}}{p_1} Z_B |B_{00}\rangle\langle B_{00}| Z_B + \frac{p_{XI} + p_{YI}}{p_1} |B_{00}\rangle\langle B_{00}|$$

is obtained with probability p_1/p_{odd} , where $p_1 = p_{XZ} + p_{YZ} + p_{XI} + p_{YI}$ and then further passed onto the hashing protocol.

For a detailed discussion of the protocol steps, see Appendix B 2. The overall (normalized per Bell pair) yield of the Ref. [10] protocol is

$$\begin{aligned}\text{Yield}_{VV}(\rho_2) &= p_{\text{even}} \left[1 - \frac{S(\rho_{\text{even}})}{2} \right] - \frac{h_b(p_{\text{even}})}{2} \\ &\quad + \frac{p_0}{2} [1 - S(\rho_{\text{odd},0})] + \frac{p_1}{2} [1 - S(\rho_{\text{odd},1})].\end{aligned}\quad (14)$$

In the case of ρ_2 being the output of independent and identically distributed channel uses, where $\rho_2 = \rho^{\otimes 2}$, with ρ [Eq. (10)] the result of sending one half of a Bell state through a Pauli noise channel, the yield expression given by Eq. (14) reduces to Eqs. (33) and (34) of Ref. [10], where it becomes apparent that the protocol indeed outperforms hashing [Eq. (12)], as

$$\begin{aligned}\text{Yield}_{VV}(\rho^{\otimes 2}) &= Y_{\text{hashing}} + \frac{p_{\text{odd}}}{4} \left(h_b \left(\frac{p_I}{p_I + p_Z} \right) \right. \\ &\quad \left. + h_b \left(\frac{p_X}{p_X + p_Y} \right) \right).\end{aligned}\quad (15)$$

The protocol of Ref. [10] was further analyzed and generalized into a scheme with improved rates in Ref. [16]. It is

The protocol consists of a step of partial breeding that uses many random infinite-size AEM checks. The check determines if the Pauli error on two consecutive Bell pairs is in $\mathcal{P}_{\text{even}} = \{IZ, ZI, XX, YY, ZZ, II, XY, YX\}$ or $\mathcal{P}_{\text{odd}} = \{XZ, ZX, XI, YI, IX, IY, YZ, ZY\}$. This step projects the state of the i th and $(i+1)$ th noisy Bell pairs with initial state ρ_2 [setting $n = 2$ in Eq. (11)] into

important to note that to achieve nonzero rates in the high-noise regimes, both the protocols presented in Refs. [10,16] still require preprocessing via a number of iterations of the recurrence protocol.

4. Adaptive entanglement purification protocol AEPP*(4)

In Ref. [15], the authors propose an adaptive protocol that relies on performing a BPM (finite-size) checks on m noisy Bell pairs to determine if the Pauli error affecting the state commutes or anticommutes with $Z^{\otimes m}$. In case the error commutes with $Z^{\otimes m}$, the state is passed onto the hashing protocol. Otherwise, further shorter finite-size checks are applied to localize the error.

A modified version of this protocol is also suggested in the same work [15], where instead of switching directly to hashing when the error commutes with $Z^{\otimes m}$, a check in the complementary basis is performed first. The authors use these ideas to devise the AEPP*(4) protocol. Under the assumption that the input states are i.i.d., the protocol proceeds as follows: First a ZZZZ check is measured over four noisy Bell pairs, giving the syndrome bit b_1 . If the outcome is 0 (indicating the error commutes with ZZZZ), then we measure XXXX, giving the syndrome bit b_2 . While the actual measurement implemented is that of XXX on the three surviving pairs, as a result of having implemented the circuit for the ZZZZ measurement, we are effectively measuring if the original error before the ZZZZ check commutes with the four-qubit check XXXX.

If the XXXX measurement yields the outcome 1 (indicating the error anticommutes with XXXX), the state is undistillable (as can be checked via the condition for Bell-diagonal states being positive partial transpose (PPT) [Eq. (A2)], discussed in Ref. [30] and restated in Appendix A 4) and is thus discarded. If the XXXX measurement yields the outcome 0, then the two-qubit state that remains is passed onto other protocols. The qubits in this state have correlated errors.

Finally, if the outcome of the ZZZZ check is 1, instead of measuring XXXX, we measure ZZ on the first two qubits, giving the syndrome bit b'_2 . If the outcome is 0, the first pair is accepted and the third is discarded. If the outcome is 1, then the third pair is discarded, while the first is kept. The accepted state is then passed onto other protocols.

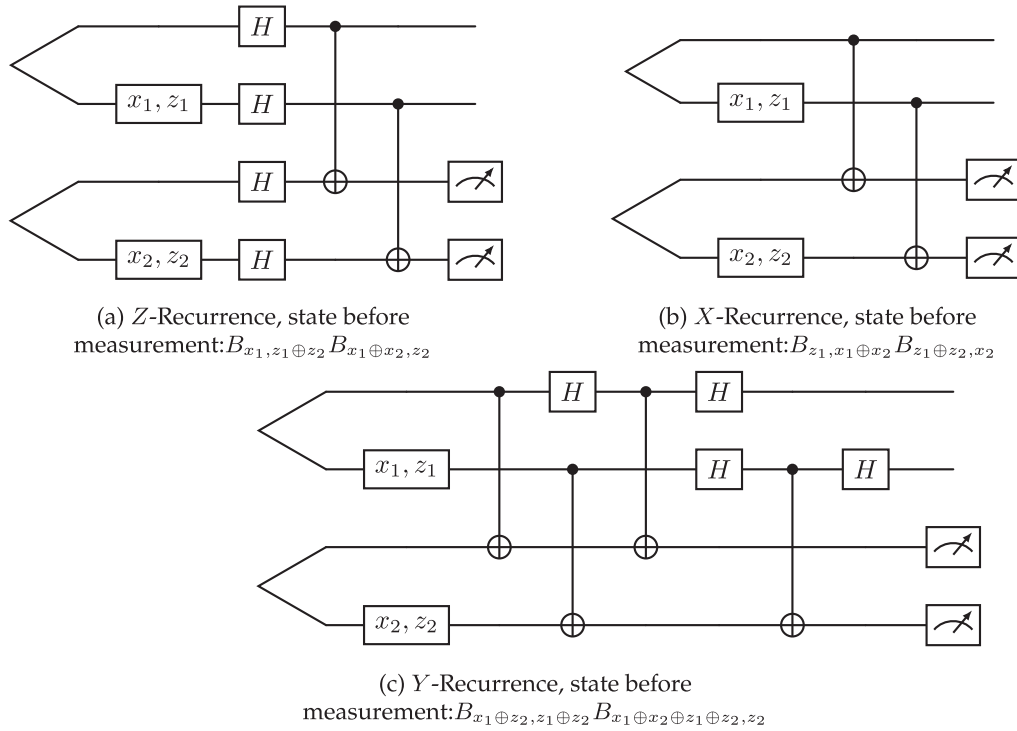


FIG. 7. Circuits for implementing ZZ, XX, and YY-check measurements.

C. Proposed combined protocol

We propose a combined protocol, which at high depolarizing noise rates, achieves higher rates than the aforementioned protocols, such as the protocol of Ref. [10] preceded by a number of recurrence iterations or the protocol of Ref. [15].

The proposed protocol can be generally summarized as follows: A number of recurrence iterations are carried out, then the accepted states are Bell states that are acted on by an effective channel that acts as i.i.d. uses of a Pauli channel that is improved in comparison to the original channel acting on the initial states. The accepted states can then either be directly passed onto the protocol of Ref. [10] or to a round of the AEPP*(4) protocol. The accepted states of the AEPP*(4) protocol are passed onto the protocol of Ref. [10].

Since our protocol is composed of all these building blocks, in the following sections, we scrutinize each component, to ensure that the decisions made at each step achieve the largest yield possible. We study the recurrence protocol, and introduce the Greedy recurrence protocol. Moreover, we study the case when AEPP*(4) protocol outputs correlated states over two Bell pairs rather than i.i.d. states. Since these correlated states are then fed into the Ref. [10] protocol, we discuss the conditions under which the Ref. [10] protocol applied to more general states still provides an advantage over hashing.

1. Further analysis of recurrence

Generally, the recurrence parity check may be a Z check, where Alice and Bob determine if the two-qubit Pauli error commutes with the stabilizer $Z \otimes Z$, or it can be an X check measuring the stabilizer $X \otimes X$ or a Y check measuring $Y \otimes Y$. Figure 7 shows circuits for realizing these measurements. After applying one of these circuits, Alice and Bob

use classical communication to compare the results of their Z-basis measurements. They only accept pairs where their measurement outcomes are equal. After one recurrence iteration, the rate is halved, since out of every two input noisy Bell pairs, at most one Bell pair survives. One may think of the recurrence measurement as a stabilizer code with a single stabilizer, $n = 2, k = 1$. These circuits impose a certain choice of logical operators and create an induced Pauli channel on the accepted pairs.

In what follows, we assume that the input noisy Bell pairs to the recurrence protocol are the result of sending halves of perfect Bell pairs through independent uses of identically distributed Pauli noise channels. Let $Q_0 \equiv X$, $Q_1 \equiv Y$, and $Q_2 \equiv Z$, with the index i of Q_i defined modulo 3. Then, the output of Q_i recurrence is accepted with probability $p_{\text{pass}}(Q_i) = (p_I + p_{Q_i})^2 + (p_{Q_{i+1}} + p_{Q_{i+2}})^2$. The accepted output state in each of the circuits of Fig. 7 after a round of Q_i recurrence will be a Bell-diagonal state with the following Pauli error probabilities:

$$\left[p'_I = \frac{p_I^2 + p_{Q_i}^2}{p_{\text{pass}}(Q_i)}, p'_X = \frac{p_{Q_{i+1}}^2 + p_{Q_{i+2}}^2}{p_{\text{pass}}(Q_i)}, \right. \\ \left. p'_{Q_{2-(i \bmod 2)}} = \frac{2p_I p_{Q_i}}{p_{\text{pass}}(Q_i)}, p'_{Q_{1+(i \bmod 2)}} = \frac{2p_{Q_{i+1}} p_{Q_{i+2}}}{p_{\text{pass}}(Q_i)} \right]. \quad (16)$$

The rejected output state for any Q_i recurrence will be a Bell-diagonal state $\rho_{AB} = p'_I |B_{00}\rangle\langle B_{00}| + p'_X X_B |B_{00}\rangle\langle B_{00}| X_B + p'_Y Y_B |B_{00}\rangle\langle B_{00}| Y_B + p'_Z Z_B |B_{00}\rangle\langle B_{00}| Z_B$, with $p'_I = p_X^*$ and $p'_Y = p_Z^*$. Let

$$q_1 = (p_I^* + p_Z^*)/2, \quad q_2 = (p_X^* + p_Y^*)/2, \\ q_3 = (p_I^* - p_Z^*)/2, \quad q_4 = (p_X^* - p_Y^*)/2.$$

TABLE I. $e^{-i\frac{\pi}{4}Q}Pe^{-i\frac{\pi}{4}Q^T}$.

$P \backslash Q$	$\mathbb{I}$	X	Y	Z
$\mathbb{I}$	$\mathbb{I}$	X	$\mathbb{I}$	Z
X	X	$\mathbb{I}$	Z	X
Y	Y	Y	Y	Y
Z	Z	Z	X	$\mathbb{I}$

The partial transpose of the rejected state is given by

$$\rho_{AB}^{T_B} = \begin{bmatrix} q_1 & 0 & 0 & q_4 \\ 0 & q_2 & q_3 & 0 \\ 0 & q_3 & q_2 & 0 \\ q_4 & 0 & 0 & q_1 \end{bmatrix}. \quad (17)$$

Since the rejected state has $p_I^* = p_X^*$ and $p_Y^* = p_Z^*$, we get $q_1 = q_2$ and $q_3 = q_4$, implying that $\rho_{AB}^{T_B} = q_1 \mathbb{I} + q_3 X \otimes X$. Thus, the eigenvalues of $\rho_{AB}^{T_B}$ are $p_I^*, p_I^*, p_Z^*, p_Z^*$. All of the eigenvalues are non-negative, thus the rejected state ρ_{AB} is PPT and therefore undistillable. This is a useful guarantee that all states discarded by the recurrence protocol, in the i.i.d. Pauli noise case, are indeed undistillable on their own.

2. Greedy recurrence

The essence of the recurrence protocol proposed in Ref. [14] is to apply rounds of parity checks, where the accepted states of one stage are passed onto a recurrence check in the next stage, such that effective Pauli channel improves gradually. However, only applying Z-recurrence checks will not necessarily produce a guaranteed improvement. For example, one round of Z recurrence applied to the Pauli channel ($p_I = 0.7$, $p_Z = 0.3$, and $p_X = p_Y = 0$) would give a channel with worse F_e of $p'_I = 0.58$. In Ref. [14], the authors describe a recurrence protocol with a modification suggested by Macchiavello, where rounds of Z recurrence are carried out, with each round being followed by a bilateral B_x rotation. Note that applying a bilateral rotation B_x followed by a round of Z recurrence, then applying a second bilateral rotation B_x to the accepted state, results in a state with a distribution as that produced by a round of Y recurrence. This can be seen by applying the permutation of the input Pauli noise vector according to Table I in Appendix A 2 to the probability distribution (16), then applying the same permutation to the output vector. The bilateral rotation B_x exchanges p_I with p_X . Thus, overall, the Macchiavello recurrence is equivalent to interleaving rounds of Z and Y recurrence.

In what follows, we introduce a more structured and justified method for choosing which type of Q recurrence is to be carried out at any stage of the recurrence procedure, called *Greedy recurrence*.

Greedy recurrence applies Q recurrence whenever the input Pauli channel has p_Q as the smallest component, $Q \in \{X, Y, Z\}$. At each stage, we check if a higher yield can be obtained by directly switching over to the protocol of Ref. [10] than that obtained by more rounds of recurrence first.

The Greedy recurrence protocol is well motivated by the following theorems that show that it provides a guaranteed

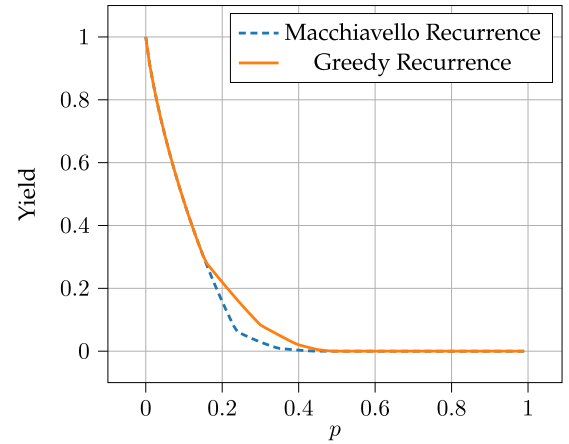


FIG. 8. Comparison of the yields of the Macchiavello recurrence and Greedy recurrence in the case of the XZ channel.

improvement of the entanglement fidelity at each step, as well as providing the largest improvement in entanglement fidelity among all choices of Q -recurrence steps.

Theorem 1. For any Pauli channel with Pauli error probability $[p_I, p_X, p_Y, p_Z]$, with $1/2 < p_I < 1$, the accepted state after one step of Greedy recurrence has a strict fidelity improvement with $p'_I > p_I$.

Theorem 2. (Optimality of Greedy recurrence for increasing fidelity) Let the input to a recurrence step be the output of two i.i.d. uses of a Pauli noise channel to send two halves of perfect Bell pairs from Alice to Bob, yielding the shared state $(\sum_{Q \in \{I, X, Y, Z\}} p_Q Q_B |B_{00}\rangle \langle B_{00}|_{AB} Q_B)^{\otimes 2}$. Assume that $p_I > 1/2$ and $p_{Q^*} = \min_{Q \in \{I, X, Y, Z\}} p_Q$. Let p'_{I, Q^*} be the entanglement fidelity after applying Q^* recurrence, while $p'_{I, \tilde{Q}}$ the entanglement fidelity after applying $\tilde{Q}$ recurrence, for $\tilde{Q} \neq Q^* \in \{X, Y, Z\}$. Then, $p'_{I, Q^*} \geq p'_{I, \tilde{Q}}$.

The proofs of Theorems 1 and 2 are provided in Appendix D.

Greedy recurrence applied to the depolarizing channel. We show that Greedy recurrence and Macchiavello recurrence apply equivalent operations (up to a final Bilateral rotation) in the case of the depolarizing channel.

Theorem 3. For the depolarizing channel with Pauli error probability $[1 - p, p/3, p/3, p/3]$, with $0 < p < 1/2$, the sequence of alternating recurrence checks $Z, Y, Z, Y, \dots$, where Z recurrence is performed, then Y recurrence is performed, and so on, is a valid sequence for Greedy recurrence.

The proof is provided in Appendix D.

Greedy recurrence applied to the XZ channel. While Greedy recurrence reproduces the Macchiavello recurrence in the case of the depolarizing channel, in the case of the XZ channel of equation [Eq. (18)], Greedy recurrence is a more useful heuristic.

$$\rho \rightarrow (1 - p)\rho + \frac{p}{2}(X\rho X + Z\rho Z). \quad (18)$$

In Fig. 8, we plot the yield obtained by applying Greedy recurrence followed by the protocol of Ref. [10], and compare it to the yield of applying the Macchiavello recurrence, again followed by the protocol of Ref. [10]. From the plot, it is clear

that Greedy recurrence provides a more general heuristic and better yields in the case of the XZ channel.

3. Further analysis of the interpolation of recurrence and hashing protocol [10]

It is clear from Eq. (15), that the yield of the protocol [10] is higher than that of hashing when the input to the protocol is the output of independent and identically distributed uses of a Pauli noise channel. However, in our proposed protocol, the input to the Ref. [10] in some cases is a correlated state of two shared Bell-diagonal pairs. In the following theorem, we verify that for permutation invariant states over the qubits held by Bob of a four-qubit shared Bell-diagonal state, such as the output of the AEPP*(4) protocol when the syndrome bits $b_1 = b_2 = 0$, the Ref. [10] gives an advantage over hashing.

Theorem 4. Given a general two-qubit Bell-diagonal state,

$$\rho_2 = \sum_{Q_1, Q_2 \in \mathcal{P}^{\otimes 2}} p_{Q_1, Q_2} Q_1 \otimes Q_2 |B_{00}\rangle \langle B_{00}|^{\otimes 2} Q_1 \otimes Q_2,$$

with $\{p_{Q_1, Q_2}\}$ a probability distribution over two-qubit Paulis. If $\{p_{Q_1, Q_2}\}$ is invariant under permutation of Q_1 and Q_2 , then $\text{Yield}_{VV}(\rho_2) \geq \text{Yield}_{\text{Hashing}}(\rho_2)$.

Proof. Using Lemma C3 twice,

$$S(\rho_2) = p_{\text{even}} S(\rho_{\text{even}}) + p_{\text{odd}} S(\rho_{\text{odd}}) + h_b(p_{\text{even}}), \quad (19)$$

$$= p_{\text{even}} S(\rho_{\text{even}}) + p_0 h\left(\frac{[p_{ZX}, p_{IX}, p_{ZY}, p_{IY}]}{p_0}\right) + p_1 h\left(\frac{[p_{XZ}, p_{XI}, p_{YZ}, p_{YI}]}{p_1}\right), \quad (20)$$

$$+ p_{\text{odd}} h_b\left(\frac{p_0}{p_{\text{odd}}}\right) + h_b(p_{\text{even}}). \quad (21)$$

For a permutation-invariant distribution, $p_0 = p_1$ and $h_b(p_0/p_{\text{odd}}) = 1$. Thus, the yield of the hashing protocol applied to ρ_2 normalized per pair is

$$\begin{aligned} \text{Yield}_{\text{Hashing}}(\rho_2) &= 1 - S(\rho_2)/2 \\ &= p_{\text{even}} + p_0/2 + p_1/2 + p_{\text{odd}}/2 - S(\rho_2)/2 \\ &= p_{\text{even}}[1 - S(\rho_{\text{even}})/2] - h_b(p_{\text{even}})/2 \\ &\quad + \frac{p_0}{2} \left[1 - h\left(\frac{[p_{ZX}, p_{IX}, p_{ZY}, p_{IY}]}{p_0}\right) \right] \\ &\quad + \frac{p_1}{2} \left[1 - h\left(\frac{[p_{XZ}, p_{XI}, p_{YZ}, p_{YI}]}{p_1}\right) \right]. \end{aligned}$$

The first equality is due to the fact that $p_{\text{even}} + p_{\text{odd}} = 1$ and $p_0 + p_1 = p_{\text{odd}}$, while the second equality is obtained by substituting $S(\rho_2)$ using Eq. (21). Then, the difference of the rates achieved by the schemes is

$$\text{Yield}_{VV}(\rho_2) - \text{Yield}_{\text{Hashing}}(\rho_2) = \frac{p_0}{2} \Delta_0 + \frac{p_1}{2} \Delta_1,$$

where

$$\begin{aligned} \Delta_0 &= h\left(\frac{[p_{ZX}, p_{IX}, p_{ZY}, p_{IY}]}{p_0}\right) - S(\rho_{\text{odd}, 0}) \\ &\stackrel{(\dagger)}{=} \frac{p_{ZX} + p_{IX}}{p_0} h_b\left(\frac{p_{ZX}}{p_{ZX} + p_{IX}}\right) \\ &\quad + \frac{p_{ZY} + p_{IY}}{p_0} h_b\left(\frac{p_{ZY}}{p_{ZY} + p_{IY}}\right), \end{aligned}$$

$$\begin{aligned} \Delta_1 &= h\left(\frac{[p_{XZ}, p_{XI}, p_{YZ}, p_{YI}]}{p_1}\right) - S(\rho_{\text{odd}, 1}) \\ &\stackrel{(*)}{=} \frac{p_{XZ} + p_{YZ}}{p_1} h_b\left(\frac{p_{XZ}}{p_{XZ} + p_{YZ}}\right) \\ &\quad + \frac{p_{XI} + p_{YI}}{p_1} h_b\left(\frac{p_{XI}}{p_{XI} + p_{YI}}\right), \end{aligned}$$

where the equalities $(\dagger)$ and $(*)$ follow by using Lemma C1. Since $\Delta_0, \Delta_1 \geq 0$, $\text{Yield}_{VV}(\rho_2) \geq \text{Yield}_{\text{Hashing}}(\rho_2)$. In this case, we may express $\text{Yield}_{VV}(\rho_2)$

$$\text{Yield}_{VV}(\rho_2) = 1 - S(\rho_2)/2 + \frac{p_{\text{odd}}}{4} (\Delta_0 + \Delta_1).$$

Further optimization. We observe that the expression in Eq. (15) gives different values for different permutations of the probability vector $[p_I, p_X, p_Y, p_Z]$. This can be exploited, by applying bilateral rotations and local Pauli operations to permute the vector to get the highest possible yield.

4. Description of the proposed protocol and results

Here, we describe the steps of the proposed protocol that is built as a careful composition of the aforementioned protocols.

(1) The input is n shared noisy Bell pairs that are generated as a result of sending n halves of Bell pairs across n i.i.d. uses of a Pauli channel.

(2) Alice and Bob apply Greedy recurrence for N_1 iterations. At each iteration, Q recurrence is applied, with Q corresponding to the Pauli error with the smallest probability in the effective channel acting on the shared states. At each iteration, the rate is halved, and the possible yield is scaled by $p_{\text{pass}}/2$, while the effective Pauli channel improves, as more errors are caught.

(3) The accepted states in the previous step are passed onto one of the two following possible branches, depending on which gives a higher yield:

(a) One choice is to apply bilateral rotations to the i.i.d. Bell pairs, to maximize the yield, then pass on these pairs to the Ref. [10] protocol, achieving a yield that can be computed via Eq. (15).

(b) The other choice is to apply the AEPP*(4) protocol. The accepted states from the Greedy recurrence step are grouped into groups of four, over which the check ZZZZ is measured.

(i) States with errors that commute with ZZZZ are subjected to the check XXXX, and if the error commutes with this check too, they are then passed onto the Ref. [10] protocol. At this point, these are correlated states over two shared Bell-diagonal pairs. These correlations are useful, since the joint entropy of a probability distribution is smaller than the sum of entropies of the marginals: $H(U_1, U_2) \leq H(U_1) + H(U_2)$. To take these correlations into consideration, we compute the yield of applying the Ref. [10] protocol via the more general expression of Eq. (14). Note that the state over the two Bell pairs at this stage is permutation invariant with respect to Bob's qubits, and by Theorem 4, the Ref. [10] protocol gives a better yield than hashing in this case.

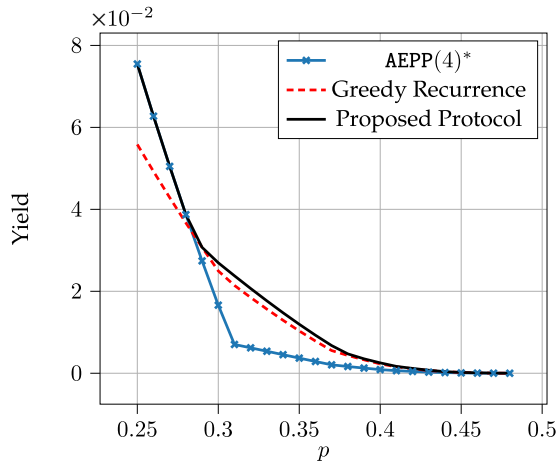


FIG. 9. Comparison of the yields of Greedy recurrence +[10], AEPP(4)*+[10], and the proposed protocol in the case of the depolarizing channel.

(ii) States that fail the ZZZZ check are subjected to a further check ZZ. If the check is passed, then the first Bell pair is passed onto a further N_2 iterations of Greedy recurrence, then to the Ref. [10] protocol, achieving (15). If the ZZ check is failed, then the third Bell pair is passed onto a further N_3 iterations of Greedy recurrence, then to the Ref. [10] protocol, achieving a yield computable by Eq. (15).

The numbers of iterations N_1 , N_2 , and N_3 can be optimized to achieve the best rate possible. In Fig. 9, we see an improved rate advantage of our proposed protocol over the protocols from the literature [10,15] for the depolarizing channel $\rho \rightarrow (1-p)\rho + p/3(X\rho X + Y\rho Y + Z\rho Z)$. Note that the AEPP(4)* protocol already exhibits an advantage over (recurrence followed by Ref. [10]) in the intermediate noise regime, which can also be seen in the plots of Ref. [15]. Intuitively, this can be explained as follows, a Z-recurrence step on four Bell pairs would catch errors such as $XIIX$, while the first check of the AEPP(4)* protocol is ZZZZ and would commute with such an error. However, both protocols would catch any single error on four pairs. After applying the checks ZZZZ, XXXX of the AEPP(4)* protocol, the rate is halved, but single-qubit errors in both amplitude and phase basis are caught. At the same cost of halving the rate, the Z-recurrence protocols only checks for amplitude errors but manages to catch more of them. Thus, we see that the AEPP(4)* protocol has a better yield than recurrence in the intermediate noise regime but not in the high-noise regime.

Our proposed protocol succeeds in extending the range of this advantage of the AEPP(4)* protocol into the high-noise regime, by preprocessing the highly noisy Pauli channel using recurrence, into a noise regime where the AEPP(4)* protocol has advantage once again.

We observe that despite the yield improvements attained by our proposed protocol, the yield per channel use is on the order of 10^{-2} . We must note, however, that we study the performance in a particularly high-noise range, where $p \geq 0.25$. In this range, the depolarizing channel is antidegradable, with zero one-way capacity. Operating in this high-noise regime in-

curs a high overhead, where we employ concatenated rounds of finite parity checks reducing the rate.

IV. DISCUSSION AND FUTURE DIRECTIONS

In this work, we have provided important examples of rate improvements for two-way assisted entanglement distillation. For the qubit depolarizing channel and other qubit Pauli channels, our techniques give higher rates for entanglement distillation than the best known well-established rates found previously. For the qubit amplitude damping channel, our techniques give higher rates than the channel's reverse coherent information, a rate that has been the best known for over a decade. Moreover, the finite preprocessing protocols and codes we provide for the amplitude damping channel are applicable to practical nonasymptotic settings. These results open avenues for investigating better protocols for distillation across a variety of channels.

We have specifically focused on the high-noise regime, where all asymptotic protocols require finite-size preprocessing parity checks to be carried out at the beginning of the protocol. Optimizing the size of the initial check used prior to an asymptotic protocol is an interesting open problem. For amplitude damping noise and Hamming-weight-1 codes, we have shown that the optimal check size is 3.

Our improvements rely on shaping multiple uses of a noisy channel into an improved channel, and carefully studying how to concatenate different distillation protocols. At each step of these protocols, we choose a distillation procedure that is most suited to the effective noise channel shaped by the previous stage. Further investigation of connections between channel reshaping (a concept akin to channel simulation [31] but assisted by two-way classical communication) and entanglement distillation may yield more insights.

In the channel reshaping procedure, the effective channel can differ from the constituent channel in two different ways. First, the reshaped channel need not belong to the same channel family as the constituent channel (see Sec. II where several amplitude damping channels are reshaped into an erasure channel). This feature can simplify the analysis if the properties of the reshaped channel are well known. Second, a single use of the reshaped channel may have higher dimensions than a single use of the constituent channel. The noise model may result in correlations between the constituent dimensions and thus may allow for additional (conceptual) freedom for improving rates for entanglement distillation over the constituent channel.

One such concept is that the joint distribution in the case of correlated noise has less entropy than the sum of entropies of the marginals. Thus, while it might be more analytically difficult to deal with distributions that are non-i.i.d., there might be gains from considering distillation protocols that yield correlated output states. One direction to circumvent the difficulty of studying general distributions might be to focus on protocols that generate correlated states with a high degree of symmetry.

ACKNOWLEDGMENTS

D.A. is grateful to IBM T.J. Watson Research Center's hospitality, where part of this work was carried out as part

of an internship. T.J.-O'.C., V.S., and J.S. were supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Co-design Center for Quantum Advantage (C2QA) Contract No. DE-SC0012704.

DATA AVAILABILITY

The data are available from the authors upon reasonable request.

APPENDIX A: USEFUL PROPERTIES OF BELL STATES

1. Ricochet property

A useful property of Bell states is the Ricochet property (or the transpose trick) [32].

Lemma A1. (Ricochet property) For any $A \in \mathbb{C}^{d \times d}$, and an orthonormal basis $\{|x\rangle\}$ of $\mathbb{C}^d$,

$$A \otimes I \sum |xx\rangle = I \otimes A^T \sum |xx\rangle,$$

where T denotes the transpose of a matrix.

Proof. We can express the operator A as $A = \sum_{u,w} a_{u,w} |u\rangle\langle w|$. The left-hand side can now be written as $\sum_{u,x} a_{u,x} |ux\rangle$.

The right-hand side is $\sum_{u,w} a_{u,w} I|x\rangle \otimes |w\rangle\langle u|x\rangle = \sum_{x,w} a_{x,w} |xw\rangle$, which equals the left-hand side by renaming the variables.

2. Bilateral rotations

We review the bilateral rotations described in Ref. [14] and how Bell states transform under such rotations. Let $R_x = e^{-i\frac{\pi}{4}X}$, $R_y = e^{-i\frac{\pi}{4}Y}$, $R_z = e^{-i\frac{\pi}{4}Z}$, and generally $R_Q = e^{-i\frac{\pi}{4}Q} = \frac{1}{\sqrt{2}}(I - iQ)$ for $Q \in \{X, Y, Z\}$. A bilateral rotation is applied on both Alice and Bob's particles: $B_x = R_x \otimes R_x$, $B_y = R_y \otimes R_y$, and $B_z = R_z \otimes R_z$. Using Lemma A1, we can observe the effect of carrying out such a bilateral rotation on a Bell state affected by a Pauli error P as follows:

$$\begin{aligned} (R_Q \otimes R_Q)(\mathbb{I} \otimes P)|B_{00}\rangle\langle B_{00}|(\mathbb{I} \otimes P)(R_Q^\dagger \otimes R_Q^\dagger) \\ = (\mathbb{I} \otimes e^{-i\frac{\pi}{4}Q} P e^{i\frac{\pi}{4}Q})|B_{00}\rangle\langle B_{00}|(\mathbb{I} \otimes e^{i\frac{\pi}{4}Q} P e^{i\frac{\pi}{4}Q}). \end{aligned}$$

The bilateral rotation then transforms the state according to Table I.

Consider the state

$$\begin{aligned} \rho = p_I |B_{00}\rangle\langle B_{00}| + p_X X_B |B_{00}\rangle\langle B_{00}| X_B + p_Y Y_B |B_{00}\rangle\langle B_{00}| Y_B \\ + p_Z Z_B |B_{00}\rangle\langle B_{00}| Z_B \end{aligned} \quad (\text{A1})$$

that results from sending half of a Bell state $|B_{00}\rangle\langle B_{00}|$ across a Pauli channel. Applying bilateral rotations to such a state permutes the Pauli operators. For example, applying B_x to ρ , we get $B_x \rho B_x^\dagger = p_I X_B |B_{00}\rangle\langle B_{00}| X_B + p_X |B_{00}\rangle\langle B_{00}| + p_Y Y_B |B_{00}\rangle\langle B_{00}| Y_B + p_Z Z_B |B_{00}\rangle\langle B_{00}| Z_B$. Thus, the channel vector is permuted from $[p_I, p_X, p_Y, p_Z]$ to $[p_X, p_I, p_Y, p_Z]$.

3. Symplectic representation

A Pauli string of length n is a tensor product of Pauli matrices acting on n qubits. Up to a global phase $(\pm i, \pm 1)$,

such a Pauli string can be represented by a binary string of length $2n$, where the first n bits represent the X component, such that the i th bit is 1 only if P has an X or Y in the i th position and 0 otherwise. The last n bits represent the Z component, such that the $(i+n)$ th bit is 1 only if P has an Z or Y in the i th position. In a similar manner, an n -pair Bell state is denoted by $|B_s\rangle = |B_{x_1, z_1}\rangle \otimes |B_{x_2, z_2}\rangle \cdots \otimes |B_{x_n, z_n}\rangle$, where $s = (x_1, z_1), \dots, (x_n, z_n)$ is a binary string of length $2n$. This mapping between Paulis and binary strings is called the symplectic representation. Let s be the symplectic representation of P and t be the symplectic representation of Q . The symplectic inner product between s and t determines whether P and Q commute or anticommute. Let s_x, t_x be the X part of the symplectic representation of P, Q , respectively, and s_z, t_z be the Z parts of P, Q , respectively. P, Q commute if and only if the symplectic inner product: $(s_x \cdot t_z + s_z \cdot t_x) \bmod 2 = 0$, else they anticommute [28]. This establishes a direct connection between linear parity checks applied to binary strings and stabilizer measurements.

4. PPT Bell-diagonal states

Here, we recap an expression for testing whether a Bell-diagonal state is PPT, as discussed in Ref. [30].

A Bell-diagonal matrix M of dimension $2^{2n} \times 2^{2n}$ is diagonal in the Bell basis:

$$M = \sum_{s \in \{0,1\}^{2n}} \alpha_s |B_s\rangle\langle B_s|.$$

Note that $B_{00}^{T_A} = \frac{1}{2}(B_{00} + B_{01} + B_{10} - B_{11})$, so

$$(B_{00}^{T_A})^{\otimes n} = \frac{1}{2^n} \sum_{k \in \{0,1\}^{2n}} (-1)^{N_{11}(k)} |B_k\rangle\langle B_k|,$$

where $N_{11}(k)$ denotes the number of occurrences of $|B_{11}\rangle$ in $|B_k\rangle$. Note that $|B_s\rangle = I_A \otimes P_{s,B} |B_{00}\rangle$, for some Pauli string P_s .

The partial transpose of M is $M^{T_A} = \frac{1}{2^n} \sum_{s,k \in \{0,1\}^{2n}} \alpha_s (-1)^{N_{11}(k)} |B_{s \oplus k}\rangle\langle B_{s \oplus k}|$. M is PPT if $\forall m \in \{0,1\}^{2n}$, $\langle B_m | M^{T_A} | B_m \rangle \geq 0$, or equivalently,

$$\forall m \in \{0,1\}^{2n}, \sum_{s \in \{0,1\}^{2n}} \alpha_s (-1)^{N_{11}(s \oplus m)} \geq 0. \quad (\text{A2})$$

5. Useful definitions

Define a commutation function as follows:

$$c(P_1, P_2) = \begin{cases} 1, & \text{if } P_1, P_2 \text{ commute,} \\ -1, & \text{otherwise.} \end{cases}$$

APPENDIX B: EXTENDED LITERATURE REVIEW

1. Parity checks flavors: AEM and BPM

a. Appended ebit measurements [16]

In the case of AEMs, Alice and Bob share an extra noiseless Bell pair. We may assume that they share a state of the form

$$\sum_{s \in \{0,1\}^{2n}} p_s |B_s\rangle\langle B_s| \otimes |B_{00}\rangle\langle B_{00}|,$$

where the first $2n$ qubits are the shared noisy Bell pairs at the output of the Pauli channel, and the final noiseless pair is the appended ebit. Let $f_x(s) = (x \cdot s) \bmod 2$, for some $x \in \{0, 1\}^{2n}$, be any linear binary parity check. The value of this check can be determined in two steps.

(1) By carrying out local Cliffords [bilateral xor; Fig. 6(a)], local Paulis, and bilateral rotations (Appendix A 2) [14], Alice and Bob can map their shared state to

$$\sum_{s \in \{0,1\}^{2n}} p_s |B_s\rangle \langle B_s| \otimes |B_{f_x(s),0}\rangle \langle B_{f_x(s),0}|.$$

Alice and Bob each measure their respective halves of the appended ebit in the Z basis, Alice can then share her measurement outcome with Bob.

(2) Using classical communication, Bob learns Alice's measurement outcome. By comparing Alice's outcome a with his own b , Bob can determine the value of $f_x(s)$ as $a \oplus b$. This can be seen as $f_x(s) = 0$ implies that the appended ebit is in the state $\frac{|00\rangle + |11\rangle}{\sqrt{2}}$ resulting in Alice's and Bob's measurement outcomes being equal. On the other hand, $f_x(s) = 1$ implies that the appended ebit is in the state $\frac{|01\rangle + |10\rangle}{\sqrt{2}}$ resulting in Alice's and Bob's measurement outcomes being unequal.

b. Bilateral Pauli measurements [16]

In a BPM, we also proceed similarly in two steps; the first is applying local Cliffords and measurements, and the second is classical communication.

(1) In the case of a BPM, Alice and Bob carry out the same projective measurement on their respective halves of the noisy Bell pairs. We study the effect of these measurements on each of the states $(\mathbb{I}_A \otimes P_B)|B_{00}\rangle^{\otimes n}$ in the decomposition of the noisy state (11). Let Q be a Pauli operator, then projectors $\Pi_+ = (\mathbb{I} + Q)/2$ and $\Pi_- = (\mathbb{I} - Q)/2$ form elements of this projective measurement. Let $a \in \{+1, -1\}$ be the outcome of Alice's measurement, $b \in \{+1, -1\}$ be the outcome of Bob's measurement, and $\tilde{c} = c(Q, P)$. The postmeasurement state is

$$\frac{\mathbb{I}_A + aQ_A}{2} \otimes \frac{\mathbb{I}_B + bQ_B}{2} (\mathbb{I}_A \otimes P_B)|B_{00}\rangle^{\otimes n}.$$

Let $\hat{Y}(P)$ be a function for the number of Y 's in a Pauli string, for example, $\hat{Y}(I \otimes Y \otimes X \otimes Y \otimes Z) = 2$. Using the Ricochet property of Lemma A1, and the fact that $Q^T = (-1)^{\hat{Y}(Q)}Q$, we may rewrite the postmeasurement state as

$$\mathbb{I}_A \otimes \left(\frac{1 + ab\tilde{c}(-1)^{\hat{Y}(Q)}}{2} \right) P_B \frac{\mathbb{I}_B + b\tilde{c}Q_B}{2} |B_{00}\rangle^{\otimes n}. \quad (\text{B1})$$

(2) The probability that Alice and Bob's measurement outcomes satisfy $ab = \tilde{c}(-1)^{\hat{Y}(Q)}$ is 1, so they can always deduce the value of $\tilde{c}(-1)^{\hat{Y}(Q)}$ by classically communicating their measurement outcomes.

By tracing out the A systems of the state in (B1), we get a state $\propto P(\mathbb{I}_B + b\tilde{c}Q_B)$ with 2^{n-1} nonzero eigenvalues. Thus, the Schmidt rank of the state (B1) is 2^{n-1} , implying that the measurement costs are 1 of the shared noisy Bell pairs.

Note that in contrast with the AEM case, BPMs have a "backaction" that alters the state shared between Alice and Bob. Such a change of the shared state must be accounted for in entanglement distillation protocols.

2. Interpolation of recurrence and hashing [10]

The protocol of Ref. [10] makes use of two main ideas; the first is partial breeding and the second is bilateral Pauli measurements. Partial breeding can be thought of as an interpolation of recurrence and hashing in the sense that in recurrence a finite-size parity check that consumes a Bell pair out of each block of two Bell pairs is carried out, determining the value of the check for each of the blocks, while hashing exactly determines the entire Pauli string at an entropic cost, rather than a fixed cost per check. In partial breeding, the same information as a finite-size check (such as recurrence) is determined, but since the check is done in an asymptotic manner like hashing, the cost depends on the entropy of the outcome distribution. Thus, the partial breeding step of the protocol of Ref. [10] determines asymptotically a linear parity check $s \cdot r$, where r is $n/2$ copies of $[1010]$, determining which blocks of two Bell pairs are affected by a single X error. On the other hand, the obtained state at the following stages of the protocol may have a negative breeding/hashing rate despite being distillable. To extract the leftover entanglement, Ref. [10] suggests carrying out a bilateral Pauli measurement, similar to a finite-size recurrence step. The backaction and degeneracy effects due to this finite-size check manage to squeeze out a state with less entropy and thus having a positive hashing rate.

Here, we explain in more detail the steps of the protocol of Ref. [10].

(1) The protocol starts by applying "partial breeding" where, similarly to breeding, many asymptotic (infinite-size) AEM random parity checks are applied. However, in contrast with breeding, instead of applying checks to completely identify s , only enough checks are applied to determine the values of $s_{4i} \oplus s_{4i+2}$ for all i . In other words, instead of completely determining the Pauli error P_B , partial breeding is used to find out, for every pair of noisy qubits $2k, 2k+1$, if the error $P_{2k,B} \otimes P_{2k+1,B}$ commutes or anticommutes with $Z_{2k} \otimes Z_{2k+1}$.

Let ρ_2 be the state of the noisy Bell pairs $2k$ and $2k+1$,

$$\begin{aligned} \rho_2 = & \sum_{P_{2k,B}, P_{2k+1,B} \in \mathcal{P}^{\otimes 2}} \text{Pr}[P_{2k,B} \otimes P_{2k+1,B}] P_{2k,B} \otimes P_{2k+1,B} |B_{00}\rangle \\ & \times \langle B_{00}|^{\otimes 2} P_{2k,B} \otimes P_{2k+1,B}, \end{aligned} \quad (\text{B2})$$

where $\text{Pr}[P_{2k,B} \otimes P_{2k+1,B}]$ is the probability that a certain Pauli error $P_{2k,B}, P_{2k+1,B} \in \mathcal{P}^{\otimes 2}$ occurs on Bob's parts of the $2k$ th, $(2k+1)$ th noisy Bell pairs, according to the channel model. Once the partial breeding checks are complete, Alice and Bob can determine if $P_{2k,B} \otimes P_{2k+1,B} \in \mathcal{P}_{\text{even}} = \{IZ, ZI, XX, YY, ZZ, II, XY, YX\}$. Let p_{even} denote the probability that $P_{2k,B} \otimes P_{2k+1,B} \in \mathcal{P}_{\text{even}}$. The number of noiseless Bell pairs to be consumed in the partial breeding steps, which is the number of AEM checks required to determine for all k if $P_{2k,B} \otimes P_{2k+1,B} \in \mathcal{P}_{\text{even}}$ is $h_b(p_{\text{even}})$ clean Bell pairs per two noisy pairs.

(2) For every pair of noisy qubits $2k, 2k+1$, if indeed $P_{2k,B} \otimes P_{2k+1,B} \in \mathcal{P}_{\text{even}}$, then the obtained state

$$\begin{aligned} \rho_{\text{even}} = & \frac{1}{p_{\text{even}}} \sum_{P_{1,B}, P_{2,B} \in \mathcal{P}_{\text{even}}} \text{Pr}[P_{1,B} \otimes P_{2,B}] P_{1,B} \otimes P_{2,B} |B_{00}\rangle \\ & \times \langle B_{00}|^{\otimes 2} P_{1,B} \otimes P_{2,B} \end{aligned}$$

is then fed into the hashing protocol, which gives the yield $2 - S(\rho_{\text{even}})$, per two noisy pairs.

(3) Otherwise, for pairs of Bell pairs where the two-qubit error $P_{2k,B} \otimes P_{2k+1,B}$ anticommutes with $Z_{2k} \otimes Z_{2k+1}$, $P_{2k,B} \otimes P_{2k+1,B} \in \mathcal{P}_{\text{odd}} = \{XZ, ZX, XI, YI, IX, IY, YZ, ZY\}$. This occurs with probability $p_{\text{odd}} = 1 - p_{\text{even}}$. The state in this case is

$$\rho_{\text{odd}} = \frac{1}{p_{\text{odd}}} \sum_{P_1 \otimes P_2 \in \mathcal{P}_{\text{odd}}} \Pr[P_{1,B} \otimes P_{2,B} | P_{1,B} \otimes P_{2,B} | B_{00}] \times \langle B_{00} |^{\otimes 2} P_{1,B} \otimes P_{2,B}.$$

In this case, a BPM (finite-size) check is applied to the state ρ_{odd} to determine if $P_{1,B}$ commutes with Z_1 . In this case, determining this commutation relation is not done via many random infinite-size measurements with appended ebits asymptotically, but through a finite-size check that applies a projective measurement onto the $+1$ and -1 eigenstates of Z on both Alice and Bob's particles of the first noisy pair. Such a finite-size check collapses one of the noisy Bell pairs of the state ρ_{odd} . Moreover, since the errors $P_1 \otimes P_2$ and $P_1 Z \otimes P_2$ differ by a stabilizer, they have the same effect on the state, and are mapped to the same state, leading to entropy reduction, as pointed out by Ref. [16]. After this measurement, we end up with a rank-2 state.

Given that the partial breeding step indicated that the two-qubit error anticommutes with ZZ , then

(1) $P_{1,B}$ commutes with Z_1 with probability $\frac{p_0}{p_{\text{odd}}}$, where $p_0 = p_{ZX} + p_{IX} + p_{ZY} + p_{IY}$. In this case, the state becomes

$$\rho_{\text{odd},0} = \frac{p_{ZX} + p_{IX}}{p_0} X_B |B_{00}\rangle \langle B_{00}| X_B + \frac{p_{ZY} + p_{IY}}{p_0} Y_B |B_{00}\rangle \langle B_{00}| Y_B \times \langle B_{00} | Y_B.$$

This state can then be passed onto the hashing protocol achieving a yield of $1 - S(\rho_{\text{odd},0})$.

(2) $P_{1,B}$ anticommutes with Z_1 with probability $\frac{p_1}{p_{\text{odd}}}$, where $p_1 = p_{XZ} + p_{XI} + p_{YZ} + p_{YI}$. In this case, the state becomes

$$\rho_{\text{odd},1} = \frac{p_{XZ} + p_{YZ}}{p_1} Z_B |B_{00}\rangle \langle B_{00}| Z_B + \frac{p_{XI} + p_{YI}}{p_1} |B_{00}\rangle \langle B_{00}|.$$

This state can then be passed onto the hashing protocol achieving a yield of $1 - S(\rho_{\text{odd},1})$.

In summary, the protocol starts by applying partial breeding to $n/2$ copies of a Bell diagonal state over two Bell pairs σ , which can be decomposed as $\sigma = p_{\text{even}} \rho_{\text{even}} \oplus p_{\text{odd}} \rho_{\text{odd}}$, where the direct sum symbol $\oplus$ expresses that ρ_{even} and ρ_{odd} are orthogonal. Then, the state

$$\sigma^{\otimes(n/2)} = \sum_{f \in \{\text{even}, \text{odd}\}^{n/2}} \Pr[f] \bigotimes_{f_i \in f} \rho_{f_i}.$$

After partial breeding, we learn an $n/2$ -bit string indicating which pairs have an error $\in \mathcal{P}_{\text{even}}$, and which have an error $\in \mathcal{P}_{\text{odd}}$. Once we learn f , the pairs in the even case are passed onto the hashing protocol obtaining $1 - S(\rho_{\text{even}})/2$ per input Bell pair, while those in the odd case are passed onto another

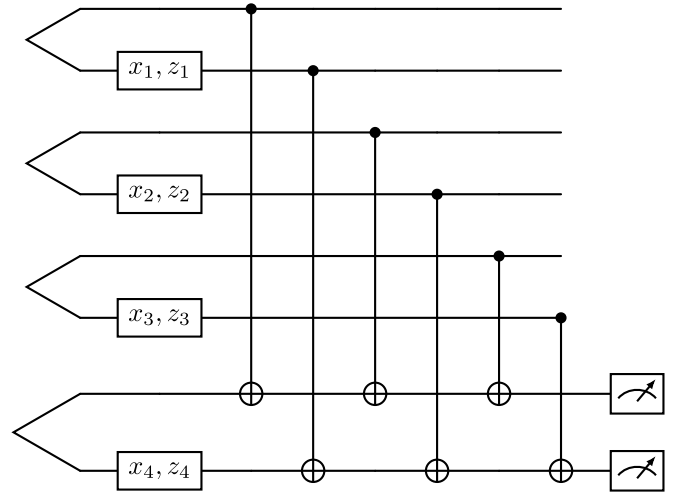


FIG. 10. ZZZZ check.

round of a BPM check, then to hashing. Since the even case occurs with probability p_{even} , and since the outcome of final BPM check is 0 with probability p_0 yielding a state $\rho_{\text{odd},0}$ and is 1 with probability $p_1 = p_{\text{odd}} - p_0 = 1 - p_{\text{even}} - p_0$ yielding a state $\rho_{\text{odd},1}$, the overall yield of the protocol is given by

$$\text{Yield}_{VV}(\rho_2) = p_{\text{even}} \left[1 - \frac{S(\rho_{\text{even}})}{2} \right] - \frac{h_b(p_{\text{even}})}{2} + \frac{p_0}{2} [1 - S(\rho_{\text{odd},0})] + \frac{p_1}{2} [1 - S(\rho_{\text{odd},1})]. \quad (\text{B3})$$

3. Adaptive entanglement purification protocol AEPP*(4)

In Ref. [15], the authors propose an adaptive protocol that relies on performing a BPM (finite-size) check on m noisy Bell pairs to determine if the Pauli error affecting the state commutes or anticommutes with $Z^{\otimes m}$. In case the error commutes with $Z^{\otimes m}$, the state is passed onto the hashing protocol. Otherwise, further shorter finite-size checks are applied to localize the error.

A modified version of this protocol is also suggested in the same work [15], where instead of switching directly to hashing when the error commutes with $Z^{\otimes m}$, a check in the complementary basis is performed first. The authors use these ideas to devise the AEPP*(4), which we review in detail.

Consider a noisy Bell-diagonal state, where four noisy Bell-pairs are shared between Alice and Bob, $\sum_{s \in \{0,1\}^8} p_s |B_s\rangle \langle B_s|$, with s being a binary string of the form $s = (x_1, z_1), (x_2, z_2), (x_3, z_3), (x_4, z_4)$.

(1) As a first step, a BPM finite-size check is carried out on four noisy Bell pairs. This check determines if the error affecting the state commutes with $Z^{\otimes 4}$. The check is carried out by applying the circuit in Fig. 10, which consists of several applications of bilateral xors. The effect of the circuit is to map s to $s' = (x_1, z_1 \oplus z_4), (x_2, z_2 \oplus z_4), (x_3, z_3 \oplus z_4), (x_1 \oplus x_2 \oplus x_3 \oplus x_4, z_4)$.

(2) If the error commutes with ZZZZ, then the syndrome bit $b_1 = x_1 \oplus x_2 \oplus x_3 \oplus x_4$ corresponding to this check takes the value $b_1 = 0$. If $b_1 = 0$, then the next step is to check for errors in the complementary basis, so we measure XXX on

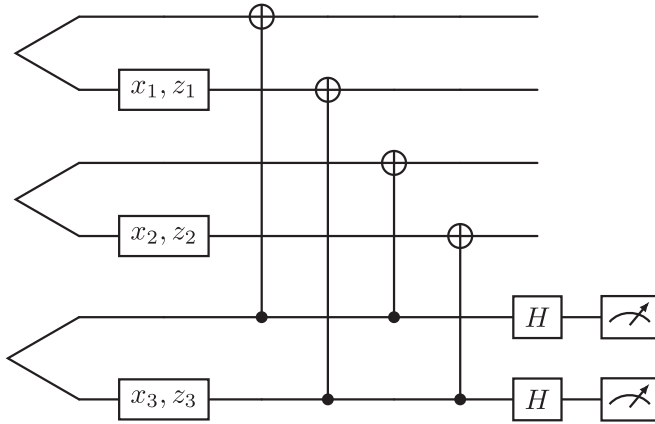


FIG. 11. XXXX check following ZZZZ check.

the state resulting from the previous check by applying the circuit in Fig. 11 to the three surviving pairs. This circuit maps the Bell string s' to $s'' = (x_1 \oplus x_3, z_1 \oplus z_4), (x_2 \oplus x_3, z_2 \oplus z_4), (z_1 \oplus z_2 \oplus z_3 \oplus z_4, x_3), (x_1 \oplus x_2 \oplus x_3 \oplus x_4, z_4)$. The syndrome bit corresponding to this check is $b_2 = z_1 \oplus z_2 \oplus z_3 \oplus z_4$. This is equivalent to having measured the XXXX check on the initial state (prior to the ZZZZ check). Note that ZZZZ, XXXX are the stabilizer checks of the well-known $[[4, 2, 2]]$ code.

After measuring the ZZZZ, XXXX checks, and obtaining the syndrome bits b_1, b_2 , the postmeasurement state, under the assumption that all four noisy Bell pairs are affected by i.i.d. Pauli noise with the probability of a Pauli error $X^x Z^z$, $x, z \in \{0, 1\}$, is proportional to

$$\sum_{r_x, r_z, t_x, t_z \in \{0, 1\}^4} q^{b_1, b_2}(r_x, r_z, t_x, t_z) |B_{r_x, r_z}\rangle \langle B_{r_x, r_z}| \otimes |B_{t_x, t_z}\rangle \langle B_{t_x, t_z}|, \quad (\text{B4})$$

where

$$\begin{aligned} q^{b_1, b_2}(r_x, r_z, t_x, t_z) &= \sum_{x_1, z_1} p(x_1, z_1) p(x_1 \oplus r_x \oplus t_x, z_1 \oplus r_z \oplus t_z) \\ &\quad \times p(x_1 \oplus r_x, b_2 \oplus z_1 \oplus t_z) \\ &\quad \times p(b_1 \oplus x_1 \oplus t_x, z_1 \oplus r_z). \end{aligned} \quad (\text{B5})$$

(3) If the error anticommutes with ZZZZ, with $b_1 = 1$, then instead of applying the XXXX check, another Z-type check is carried out to localize the error that caused the first check to fire. To do so, we measure the check ZZII. Let b'_2 be the syndrome bit obtained upon measuring ZZII, the leftover postmeasurement state (on the first and third pairs) is proportional to the following product state:

$$\begin{aligned} &\sum_{x_1, z_1, z_1 \in \{0, 1\}} p(x_1, z_1) p(x_2 = b'_2 \oplus x_1, z_2 = z_1 \oplus z_1) \\ &\quad \times |B_{x_1, z_1}\rangle \langle B_{x_1, z_1}| \\ &\otimes \sum_{x_3, z_3, z_3 \in \{0, 1\}} p(x_3, z_3) p(x_4 = 1 \oplus b'_2 \oplus x_3, z_4 = z_3 \oplus z_3) \\ &\quad \times |B_{x_3, z_3}\rangle \langle B_{x_3, z_3}|. \end{aligned}$$

When $b'_2 = 0$, the first pair is accepted pair, while the third pair becomes undistillable, and is therefore discarded. When

$b'_2 = 1$, the first pair becomes undistillable and the third pair is kept. The output state can then be passed onto other protocols. In either case, the accepted pair has a probability distribution representing an accepted pair in one step of a Z-recurrence protocol, i.e.,

$$\begin{aligned} \left[p'_I &= \frac{p_I^2 + p_Z^2}{p_{\text{accept}}}, p'_X = \frac{p_X^2 + p_Y^2}{p_{\text{accept}}}, p'_Y = \frac{2p_Y p_X}{p_{\text{accept}}}, \right. \\ p'_Z &= \left. \frac{2p_I p_Z}{p_{\text{accept}}} \right], \end{aligned} \quad (\text{B6})$$

where, $p_{\text{accept}} = (p_I + p_Z)^2 + (p_X + p_Y)^2$.

APPENDIX C: USEFUL LEMMAS

The following lemmas will be useful for analyzing different distillation protocols.

Lemma C1. (Entropy reduction by grouping) [16]

$$\begin{aligned} \forall x, y \geq 0, -x \log_2 x - y \log_2 y + (x + y) \log_2(x + y) \\ = (x + y) h_b\left(\frac{x}{x + y}\right) \geq 0. \end{aligned}$$

Lemma C2. (Inequality of ratios of arithmetic mean to geometric mean of pairs of disparate probabilities versus pairs of similar probabilities) Let a, b, c , and d be the elements of a probability distribution satisfying $a = \frac{1}{2} + \epsilon$, $\epsilon > 0$, and $a \geq c \geq d \geq b$. Then,

$$cd(a^2 + b^2) - ab(c^2 + d^2) \geq 0.$$

Proof. We may factorize the left-hand side of the inequality as $(ad - bc)(ac - bd)$. Due to the ordering of the variables $a \geq c \geq d \geq b$, we have $ad \geq bc$ and $ac \geq bd$. Thus, both factors of the left-hand side satisfy $(ad - bc) \geq 0$ and $(ac - bd) \geq 0$ proving the inequality.

Lemma C3. (Entropy of a direct sum) Let $\rho = \sum_{i \in S_0} \lambda_i |\psi_i\rangle \langle \psi_i| + \sum_{i \in S_1} \lambda_i |\psi_i\rangle \langle \psi_i|$, where S_0, S_1 are two disjoint sets of indices, $\sum_{i \in S_0 \cup S_1} \lambda_i = 1$ and $\langle \psi_k | \psi_{k'} \rangle = \delta_{kk'} \forall k, k' \in S_0 \cup S_1$. Moreover, let

$$p_x = \sum_{i \in S_x} \lambda_i, \quad \tilde{\rho}_x = \sum_{i \in S_x} \frac{\lambda_i}{p_x} |\psi_i\rangle \langle \psi_i|, x \in \{0, 1\},$$

i.e., $\rho = p_0 \tilde{\rho}_0 \oplus p_1 \tilde{\rho}_1$. Then,

$$S(\rho) = p_0 S(\tilde{\rho}_0) + p_1 S(\tilde{\rho}_1) + h_b(p_0).$$

Proof. Extend $\{|\psi_i\rangle\}$ to a full orthonormal basis, and consider the isometry

$$V = \sum_i |f(i)\rangle |\psi_i\rangle \langle \psi_i|,$$

with $f(i) = 0$ if $i \in S_0$, $f(i) = 1$ if $i \in S_1$, and $f(i) = 2$ if i not in S_0 or S_1 . Note that V satisfies $V^\dagger V = I$. Since entropy does not change under the application of an isometry $S(\rho) = S(V \rho V^\dagger)$, where $V \rho V^\dagger$ is a classical quantum state given by

$$V \rho V^\dagger = \sum_{x \in \{0, 1\}} p_x |x\rangle \langle x| \otimes \tilde{\rho}_x.$$

Due to this identity for joint entropy of a classical quantum state $S(B, X) = S(B|X) + S(X)$, we may write

$$S(\rho) = S(V\rho V^\dagger) = \sum_x p_x S(\tilde{\rho}_x) + h_b(p_0).$$

APPENDIX D: PROOFS OF THEOREMS PROVIDING PERFORMANCE GUARANTEES FOR GREEDY RECURRENCE

Theorem D1. For any Pauli channel with Pauli error probability $[p_I, p_X, p_Y, p_Z]$, with $1/2 < p_I < 1$, the accepted state after one step of Greedy recurrence has a strict fidelity improvement with $p'_I > p_I$.

Proof. Since we focus on the range of input states with $p_I > 1/2$, we may write $p_I = 1/2 + \epsilon$ for $0 < \epsilon < 1/2$.

By construction of the Greedy recurrence scheme, at each step we apply Q recurrence, with Q being the Pauli channel component with the smallest probability p_Q , i.e., $p_Q \leq p_{Q'}, \forall Q' \in \{I, X, Y, Z\}$. Then, $p_Q = (1/2 - \delta)/3$ with $\epsilon \leq \delta \leq 1/2$. Using Eq. (16), we find that the improvement is

$$p'_I - p_I = \epsilon \frac{\alpha - \gamma}{\beta + \gamma},$$

with

$$\alpha = 81 + 648\delta,$$

$$\beta = 405 - 162\delta,$$

$$\gamma = 162\delta^2 - 972\delta\epsilon + 1458\epsilon^2 + 486\epsilon.$$

For $0 < \epsilon < 1/2$, $\epsilon \leq \delta \leq 1/2$, $\beta > 0$, $\gamma > 0$. Assuming a strict inequality: $\epsilon < \delta$, we can show

$$\begin{aligned} \gamma &< 486\epsilon + 486\delta\epsilon + 162\delta^2 \\ &< 648\delta + 162\delta(4\delta - 1) \\ &\leq 648\delta + 81 = \alpha. \end{aligned}$$

The first two strict inequalities are due to the assumption that $\epsilon < \delta$, while the last inequality is due to the constraint that $\delta \leq 1/2$. The previous set of inequalities implies $\gamma < \alpha$. In the case when $\epsilon = \delta$, then the strict inequality $\delta < 1/2$ holds in the third step instead of the first and second, due to $p_I < 1$. This ensures that in both cases, $\gamma < \alpha$ and $p'_I > p_I$.

Theorem D2. (Optimality of Greedy recurrence for increasing fidelity) Let the input to a recurrence step be the output of two i.i.d. uses of a Pauli noise channel to send two halves of perfect Bell pairs from Alice to Bob, yielding the shared state $(\sum_{Q \in \{I, X, Y, Z\}} p_Q Q_B |B_{00}\rangle \langle B_{00}|_{AB} Q_B) \otimes^2$. Assume that $p_I > 1/2$ and $p_{Q^*} = \min_{Q \in \{I, X, Y, Z\}} p_Q$. Let p'_{I, Q^*} be the entanglement fidelity after applying Q^* recurrence, while $p'_{I, \tilde{Q}}$ the entanglement fidelity after applying $\tilde{Q}$ recurrence, for $\tilde{Q} \neq Q^* \in \{X, Y, Z\}$. Then, $p'_{I, Q^*} \geq p'_{I, \tilde{Q}}$.

Proof. Using Eq. (16), we have

$$p'_{I, Q^*} - p'_{I, \tilde{Q}} = \frac{p_I^2 + (\alpha - p_I)^2}{\alpha^2 + (1 - \alpha)^2} - \frac{p_I^2 + (\beta - p_I)^2}{\beta^2 + (1 - \beta)^2},$$

where $\alpha = p_I + p_{Q^*}$ and $\beta = p_I + p_{\tilde{Q}}$. To show that $p'_{I, Q^*} \geq p'_{I, \tilde{Q}}$, it suffices to show that the numerator of the above ex-

pression is non-negative, as the denominator is a product of sums of squares.

Thus, we write the numerator as

$$\begin{aligned} f &= [p_I^2 + (p_I - \alpha)^2][\beta^2 + (1 - \beta)^2] \\ &\quad - [p_I^2 + (p_I - \beta)^2][\alpha^2 + (1 - \alpha)^2]. \end{aligned}$$

Note that $p_{Q^*} \leq p_{\tilde{Q}}$, and $p_I \geq 1/2$.

Rewriting, $f = (2p_I - 1)(p_{\tilde{Q}} - p_{Q^*})(2p_I^2 + p_{\tilde{Q}} + p_{Q^*} - 2p_{\tilde{Q}}p_{Q^*}) \geq 0$ by the arithmetic-geometric mean inequality.

Proof. We prove this by induction. To fix notation, we denote by $t \in \{0, 1, \dots\}$ the t th step of recurrence, with input probabilities $\{p_Q^{(t)}\}$, and output probabilities $\{p_Q^{(t+1)}\}$, computed using Eq. (16), for $Q \in \{I, X, Y, Z\}$. The proof uses induction to show that if at the initial step p_Z is equal to the minimum probability, then at following odd numbered steps $\min_{Q \in \{I, X, Y, Z\}} p_Q^{(2t+1)} = p_Y^{(2t+1)}$ and at following even-numbered steps $\min_{Q \in \{I, X, Y, Z\}} p_Q^{(2t)} = p_Z^{(2t)}$ implying that $Z, Y, Z, Y, \dots$ is a valid sequence for Greedy recurrence.

Base case: At $t = 0$, since $p_Z^{(0)} = p_X^{(0)} = p_Y^{(0)} = p/3$, i.e., $p_Z^{(0)} \leq p_Y^{(0)} \leq p_X^{(0)} \leq p_I^{(0)}$, we perform a step of Z recurrence. At the output, we have

$$p_I^{(1)} = \frac{(1-p)^2 + p^2/9}{p_{\text{pass}_Z}}, \quad p_Z^{(1)} = \frac{2(1-p)p}{3p_{\text{pass}_Z}},$$

$$p_X^{(1)} = \frac{2p^2}{9p_{\text{pass}_Z}}, \quad p_Y^{(1)} = \frac{2p^2}{9p_{\text{pass}_Z}}.$$

At $t = 1$, since $p < 1/2$, $p_Y^{(1)} \leq \min\{p_Z^{(1)}, p_X^{(1)}\}$, we apply Y recurrence. At the output, we get $p_X^{(2)} \geq p_Z^{(2)}$ and

$$\begin{aligned} p_Z^{(2)} &= \frac{2p_X^{(1)}p_Z^{(1)}}{p_{\text{pass}_Y}} = \frac{4[2p^2/9](1-p)p/3}{p_{\text{pass}_Y}p_{\text{pass}_Z}^2} \\ &= \frac{p_Y^{(1)}4(1-p)p/3}{p_{\text{pass}_Y}p_{\text{pass}_Z}} \leq \frac{2p_Y^{(1)}p_I^{(1)}}{p_{\text{pass}_Y}} = p_Y^{(2)}. \end{aligned}$$

Inductive step:

At (even-numbered) step $2t$: assume

$$p_Z^{(2t)} \leq \min\{p_Y^{(2t)}, p_X^{(2t)}\}, \quad p_Y^{(2t-1)} \leq \min\{p_Z^{(2t-1)}, p_X^{(2t-1)}\}.$$

These inequalities imply that Y recurrence was performed at step $2t - 1$ and Z recurrence will be performed at step $2t$. Then,

$$p_{\text{pass}_Z}p_X^{(2t+1)} = (p_X^{(2t)})^2 + (p_Y^{(2t)})^2 \geq 2p_X^{(2t)}p_Y^{(2t)} = p_{\text{pass}_Z}p_Y^{(2t+1)}.$$

By Theorem D1

$$p_I^{(2t+1)} > p_I^{(2t)} > p_I^{(2t-1)} > 1/2.$$

Moreover,

$$\begin{aligned} p_{\text{pass}_Z}p_Z^{(2t+1)} &= 2p_I^{(2t)}p_Z^{(2t)} \\ &= 4 \frac{1}{p_{\text{pass}_Y}^2} [(p_I^{(2t-1)})^2 + (p_Y^{(2t-1)})^2] p_X^{(2t-1)} p_Z^{(2t-1)}. \end{aligned}$$

Using Lemma C2, with $a = p_I^{(2t-1)}$, $b = p_Y^{(2t-1)}$, $c = \max\{p_X^{(2t-1)}, p_Z^{(2t-1)}\}$, and $d = \min\{p_X^{(2t-1)}, p_Z^{(2t-1)}\}$, we have

$$\begin{aligned} p_{\text{pass}_Z}p_Z^{(2t+1)} &\geq 4 \frac{1}{p_{\text{pass}_Y}^2} p_I^{(2t-1)} p_Y^{(2t-1)} [(p_X^{(2t-1)})^2 + (p_Z^{(2t-1)})^2] \\ &= 2p_Y^{(2t)}p_X^{(2t)} = p_{\text{pass}_Y}p_Y^{(2t+1)}. \end{aligned}$$

Thus, we have shown that $p_Y^{(2t+1)} \leq \min\{p_Z^{(2t+1)}, p_X^{(2t+1)}\}$, and the next step will be a Y recurrence.

At (odd-numbered) step $2t + 1$: assume

$$p_Y^{(2t+1)} \leq \min\{p_Z^{(2t+1)}, p_X^{(2t+1)}\}, p_Z^{(2t)} \leq \min\{p_Y^{(2t)}, p_X^{(2t)}\}.$$

These inequalities imply that Y recurrence is performed at step $2t + 1$ and Z recurrence was performed at step $2t$. Then,

$$\begin{aligned} p_{\text{pass}_Y} p_X^{(2t+2)} &= (p_X^{(2t+1)})^2 + (p_Z^{(2t+1)})^2 \\ &\geq 2p_X^{(2t+1)} p_Z^{(2t+1)} = p_{\text{pass}_Y} p_Z^{(2t+2)}. \end{aligned}$$

By Theorem D1,

$$p_I^{(2t+2)} > p_I^{(2t+1)} > p_I^{(2t)} > 1/2.$$

Moreover,

$$\begin{aligned} p_{\text{pass}_Y} p_Y^{(2t+2)} &= 2p_I^{(2t+1)} p_Y^{(2t+1)} \\ &= 4 \frac{1}{p_{\text{pass}_Z}^2} [(p_I^{(2t)})^2 + (p_Z^{(2t)})^2] p_X^{(2t)} p_Y^{(2t)}. \end{aligned}$$

Using Lemma C2, with $a = p_I^{(2t)}, b = p_Z^{(2t)}, c = \max\{p_X^{(2t)}, p_Y^{(2t)}\}$, and $d = \min\{p_X^{(2t)}, p_Y^{(2t)}\}$, we have

$$\begin{aligned} p_{\text{pass}_Y} p_Y^{(2t+2)} &= 2p_I^{(2t+1)} p_Y^{(2t+1)} \\ &\geq 4 \frac{1}{p_{\text{pass}_Z}^2} p_I^{(2t)} p_Z^{(2t)} [(p_X^{(2t)})^2 + (p_Y^{(2t)})^2] \\ &= 2p_Z^{(2t+1)} p_X^{(2t+1)} = p_{\text{pass}_Y} p_Z^{(2t+2)}. \end{aligned}$$

Thus, we have shown that $p_Z^{(2t+2)} \leq \min\{p_Y^{(2t+2)}, p_X^{(2t+2)}\}$, and the next step will be a Z recurrence.

-
- [1] 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).
 - [2] C. H. Bennett and P. W. Shor, Quantum information theory, *IEEE Trans. Inf. Theory* **44**, 2724 (1998).
 - [3] H. Barnum, E. Knill, and M. A. Nielsen, On quantum fidelities and channel capacities, *IEEE Trans. Inf. Theory* **46**, 1317 (2000).
 - [4] D. Kretschmann and R. F. Werner, *Tema con variazioni*: Quantum channel capacity, *New J. Phys.* **6**, 26 (2004).
 - [5] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin, Capacities of quantum erasure channels, *Phys. Rev. Lett.* **78**, 3217 (1997).
 - [6] S. Lloyd, Capacity of the noisy quantum channel, *Phys. Rev. A* **55**, 1613 (1997).
 - [7] P. W. Shor, Quantum error correction (2002), <https://www.slmath.org/workshops/203/schedules/1181>.
 - [8] I. Devetak, The private classical capacity and quantum capacity of a quantum channel, *IEEE Trans. Inf. Theory* **51**, 44 (2005).
 - [9] I. Devetak and A. Winter, Distillation of secret key and entanglement from quantum states, *Proc. R. Soc. A* **461**, 207 (2005).
 - [10] K. G. H. Vollbrecht and F. Verstraete, Interpolation of recurrence and hashing entanglement distillation protocols, *Phys. Rev. A* **71**, 062325 (2005).
 - [11] M. Horodecki, P. Horodecki, and R. Horodecki, Unified approach to quantum capacities: Towards quantum noisy coding theorem, *Phys. Rev. Lett.* **85**, 433 (2000).
 - [12] I. Devetak, M. Junge, C. King, and M. B. Ruskai, Multiplicativity of completely bounded p-norms implies a new additivity result, *Commun. Math. Phys.* **266**, 37 (2006).
 - [13] R. García-Patrón, S. Pirandola, S. Lloyd, and J. H. Shapiro, Reverse coherent information, *Phys. Rev. Lett.* **102**, 210501 (2009).
 - [14] 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).
 - [15] A. W. Leung and P. W. Shor, Adaptive entanglement purification protocols with two-way classical communication, [arXiv:quant-ph/0702156](https://arxiv.org/abs/quant-ph/0702156).
 - [16] E. Hostens, J. Dehaene, and B. De Moor, Publisher's note: Asymptotic adaptive bipartite entanglement-distillation protocol, *Phys. Rev. A* **73**, 062337 (2006); **74**, 019903 (2006).
 - [17] M. M. Wilde, *Quantum Information Theory*, 2nd ed. (Cambridge University Press, Cambridge, 2017).
 - [18] J. Watrous, *The Theory of Quantum Information*, 1st ed. (Cambridge University Press, Cambridge, 2018).
 - [19] B. Schumacher and M. A. Nielsen, Quantum data processing and error correction, *Phys. Rev. A* **54**, 2629 (1996).
 - [20] M. M. Wolf and D. Pérez-García, Quantum capacities of channels with small environment, *Phys. Rev. A* **75**, 012303 (2007).
 - [21] I. Devetak and P. W. Shor, The capacity of a quantum channel for simultaneous transmission of classical and quantum information, *Commun. Math. Phys.* **256**, 287 (2005).
 - [22] S. Khatri, K. Sharma, and M. M. Wilde, Information-theoretic aspects of the generalized amplitude-damping channel, *Phys. Rev. A* **102**, 012401 (2020).
 - [23] R. Duan, M. Grassl, Z. Ji, and B. Zeng, Multi-error-correcting amplitude damping codes, in *2010 IEEE International Symposium on Information Theory* (IEEE, Piscataway, NJ, 2010), Vol. 75, pp. 2672–2676.
 - [24] A. Kubica, A. Haim, Y. Vaknin, H. Levine, F. Brandão, and A. Retzker, Erasure qubits: Overcoming the T_1 limit in superconducting circuits, *Phys. Rev. X* **13**, 041022 (2023).
 - [25] V. Siddhu, D. Abdelhadi, T. Jochym-O'Connor, and J. Smolin, Entanglement sharing across a damping-dephasing channel, in *2024 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Piscataway, NJ, 2024), pp. 1432–1437.
 - [26] D. Gottesman, Stabilizer codes and quantum error correction, thesis, California Institute of Technology, 1997.
 - [27] A. Ambainis and D. Gottesman, The minimum distance problem for two-way entanglement purification, *IEEE Trans. Inf. Theory* **52**, 748 (2006).

- [28] N. Rengaswamy, R. Calderbank, H. D. Pfister, and S. Kadhe, Synthesis of logical Clifford operators via symplectic geometry, in *2018 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Piscataway, NJ, 2018), pp. 791–795.
- [29] B. Schumacher, Sending quantum entanglement through noisy channels, *Phys. Rev. A* **54**, 2614 (1996).
- [30] D. P. DiVincenzo, D. W. Leung, and B. M. Terhal, Quantum data hiding, *IEEE Trans. Inf. Theory* **48**, 580 (2002).
- [31] M. B. Hastings, Notes on some questions in mathematical physics and quantum information, [arXiv:1404.4327](https://arxiv.org/abs/1404.4327).
- [32] I. Lim, Lecture notes in quantum information theory (2019), <https://lim.physics.ucdavis.edu/teaching/files/qi-notes-partiii.pdf>.